THIS BOOK IS A PART
OF THE LIBRARY OF =

A.V. LEPPANEN

THE TRVE VNIVERSITY IS A
COLLECTION OF BOOKS. CARLYLE

ELEMENTARY MATHEMATICAL STATISTICS

ELEMENTARY MATHEMATICAL STATISTICS

BY

WILLIAM DOWELL BATEN, Ph.D.

ASSOCIATE PROFESSOR OF MATHEMATICS AND RESEARCH ASSOCIATE IN STATISTICS
OF THE MICHIGAN AGRICULTURAL EXPERIMENT STATION
MICHIGAN STATE COLLEGE

NEW YORK
JOHN WILEY & SONS, INC.
LONDON: CHAPMAN & HALL, LIMITED
1938

Copyright, 1938
BY
William Dowell Baten

All Rights Reserved

This book or any part thereof must not be reproduced in any form without the written permission of the publisher.

PRINTED IN U. S. A.

PRESS OF
BRAUNWORTH & CO., INC.
BUILDERS OF BOOKS
BRIDGEPORT, CONN.

DEDICATED TO MY WIFE

ALLIE

WHO HAS ENCOURAGED ME TO PRESS
FORWARD TOWARD HIGHER GOALS

PREFACE

This book was written primarily for students interested in statistics who have not studied differential and integral calculus. It attempts to develop formulas and fundamental relations by the use of very simple algebra, trigonometry, and analytical geometry. Each bit of theory is illustrated by an example. The book contains many problems for students to solve, because the author believes that one learns statistics to a great extent by solving problems designed to bring out the meaning of theory.

Where theory requires advanced mathematics, examples are given to point out the plausibleness of the theory and to furnish information which should make the theory more acceptable. These examples should help students bridge difficult gaps between theory and application.

Comparisons are presented for distinguishing differences between the concepts of linear correlation, non-linear correlation, and correlation based on the correlation ratio. Details have not been spared in the presentation of partial, multiple, and tetrachoric correlation.

Ideas concerning sampling are developed by sampling from a finite parent population before going to the infinite; this method of approach should clarify a great deal of the theory with regard to the characteristics of distributions of averages, variances, and other statistics. This manner of introducing the most important concepts in this field should make them easier to understand and apply.

This book can be used by those interested in business and economic statistics, for it contains, in addition to needed subject matter in these fields, chapters on index numbers, trend analysis, analysis of time series, analysis of variance, and the methods of presenting data by graphs and charts.

Significance between statistics is introduced in the material pertaining to sampling and later discussed in an entire chapter.

Near the end of the book is a short chapter on the analysis of variance, which is the most recent development in applied statistics, and which is being used extensively throughout the world in

the various experimental stations and laboratories. The introduction of this rapidly developing subject should stimulate desires to do further study concerning this valuable technique for analyzing experimental data.

It is hoped that answers to the problems will make the book more usable as a textbook and as a reference book.

WILLIAM DOWELL BATEN.

MICHIGAN STATE COLLEGE
March, 1938.

CONTENTS

ERRATA

Page 19: Pop. U. S. A. 1820 = 9,638,000.
Page 34: $M_2 = 138\ 22$; $M_7 = 138.13$.
Line 29: Include 7's.
Page 60, line 11: $\sigma_2 = \sqrt{0.0625}$.
Page 61: 2nd and 4th items in last column are 2,548 and 13,365.
Page 68: Frequency for 39 is 388.
Page 70: First numbers in columns 4, 6, and 7 are 68, -340, and 1,700.
Page 73, Prob 2: last class is 113–116.
Page 74: In σ_{v+w}, 84.3 should read 74.3.
Page 83, Formula (4.2): $Y = \dfrac{1}{\sqrt{2\pi}}$, etc.
Line 16: 2nd column should be 0.0437.
Page 87: Column 4, 10th, 12th, and 20th items are 27760, 57535, and .99952.
Column 5, change accordingly.
Page 89, line 2: $M = 7.006$, S.D. = .1219.
Prob. 5: Freq. of class 64.5–69.5 is 137.
Page 99, line 1: $M_v = \mu_v' = 39$ gallons.
Page 100, line 17: $M_{3:v}$ should be $\mu_{3.v}$.
Page 106, line 29: $K = 1.8265$.
Page 117, line 16: change success to successes.
Page 119, line 7: last 3 words should read "nearest 0.1 unit."
Page 121: Frequency for 72 is 27.
Page 123: Frequency for 4 is $126(\frac{2}{3})^5(\frac{1}{3})^4$.
Page 146, line 29, second part: $9x$ instead of $9y$.
Last line should begin with -0.03.
Page 148: last line $\Sigma e_i^2 = \Sigma b_i^2 y^2$, etc.
Page 149: 3rd equation in prob. 2 is

$$-x - y + z = 0.1.$$

Third equation in prob. 3 is

$$-7x + 4y = +1.1.$$

Page 160, Formula (9.18): $b = \dfrac{\Sigma \bar{x}\bar{y}}{\Sigma \bar{x}^2}$.
Page 166, line 6: Coef. of $c = 4{,}503.83$.
Line 11: Second (2) should be (3).
Line 18: $c = 7.145$, $b = 3.106$,
$a = -120.217$.

Page 181: $r = \dfrac{8{,}117.90}{11{,}824.60} = 0.687$.
Page 184, column 3: 4th, 5th, and 11th items are 0.917, 0 866, and 0.346.
Page 188, line 5: $R = 0.936$.
Line 7: is 1 should read $R_{y\cdot xz} = 1$.
Page 191: $c_1 = -2.61$. Change predicting equation accordingly.
Page 193: last number in table is 34.8 instead of 20.2.
Page 199, line 15: Fraction is $\dfrac{55{,}612}{55{,}899.154}$.
Page 206: $\bar{z}_2^2 = [\quad]^2$.
Page 208, line 24: First S.D. is σ_z.
Page 209, line 2: $\sigma_z = \sigma_v/\sqrt{r}$.
Line 18: $M_z = M_v$.
Page 212, line 15: "the sample means shall" should read "the sample shall."
Page 221, prob. 5: Last class is 26–27.
Page 228, line 4: Insert "weighted" before the word "means."
Page 231, line 5: $r = \sqrt{0.8757}$.
Page 234, Table 13.1: Price for Jan., 1933, is 19.85.
Page 235, Table 13.2: Ave. Cor. for Nov. is 37.59.
Page 266, line 15:

$$\text{Significance} = \frac{25}{39} < 2.6.$$

Page 269, below Table 15.1:

$$^{\dagger}V = 100(\sigma/M).$$

Page 273, line 3: $\sigma_{M_y} = 1.34$.
Page 275, line 13: $t = -2.459$.
Line 14: 49 degrees should read 24 degrees.
Page 276, lines 6 and 8: $s = 0.03623$;

$$t = \frac{.0252-0}{.03623}\sqrt{\frac{25.25}{50}} = 2.459.$$

Page 277, soil C: 3rd, 6th, 8th, and 9th items are 7.0, 7.3, 7.2, and 7.2.
Line 22: Change 0.1056 to 0.12.

ANSWER CORRECTIONS

Page 5: 23. Omit last x in ( ).
Page 40: 5. (*a*) 175.
Page 48: 2. (*a*) $r = 0.0147$. 5. (*b*) 0.2418.
Page 52: 3. (*a*) 9.39. (*b*) 46 95. 8. 0.4.
Page 58: 3. (*a*) 1.52. (*b*) 56%. (*c*) 89%.
Page 65: 3. (*a*) 59.4%. 8. 4 995.
Page 72: 1. S D. = 2.606, 62%, 98.4.
4. 94.7. should read 94 1.
Page 76: 2. $M = 65\ 63$, S D. = 3.33.
Page 80: 3.(*b*) $M = 238.66$.
Page 89: 3. A's = 1.76, C's = 51.61; sum = 99.95.
Page 95: 2. $t_1 = 0.81$, area = 0 11555, $N = 3003$.
Page 102: 3. 2352 86. 6. $M = 26\ 97$, S.D = 8.7, $\alpha_3 = 1.69$ 7. M^3 instead of M^2 in 1st [ ].
Page 107: 1. Skew = − 0.48.
2. $K = 0\ 462$. 3 $\alpha_4 = 2.88$.
Page 110: 9. 1.
Page 113: 4. 720. 5. (*b*) 35. 6. (*c*) 840
12. 47.
Page 114: 2. (*b*) 40/84. 3. Omit (*d*) and (*e*). 6. (*d*) 0.06668.
Page 118: 1. (*a*) 135/4096. 7. (*a*) 0 03438. (*e*) 0.16258. 8. $w = 21\ 8$.
Page 128: 1. .2076. 2. (*a*) $252/2^{10}$, (*b*) $120/2^{10}$.
Page 143: 5. $I = 91$ (57 lb. in 1 bu. tom.).
Page 149: 6. $BC = 11.2$.
Page 153: 1. $\sigma_e = 0.73$.
Page 159: 1. $y = -\ 142{,}198\ 1 + 75.10x$; $\sigma_e = 53.47$.
Page 162: 1 $\sigma_e = 0.14$.
Page 165: 2. $\sigma_e = 8\ 39$.
3. Const. = − 120.3.
Page 167: 1. $y = -\ 202.191 + 1.477x + 3.258z + 6.052w$; $\sigma_e = 4\ 48$.
Page 175: 1. $R = 0\ 677$.
Page 182: 1. $r = 0.648$. 2. $r = 0.768$.
Page 188: $r = +\ 0.99$ or $-\ 0\ 90$.
5. $R = 0.838$.
Page 192· 4. $D_0 = .30$, $D_2 = .03$, $\bar{y} = 2.39\bar{x}_1 + .95^+\bar{x}_2 + 9.33\bar{x}_3$, $R = 0.964$.
Page 196: $r_{me\ a} = 0.0259$. 2. $r_{wh} = 0.568$, $r_{wt \cdot h} = 0.904$ 3. $r_{m\,4\,\text{yr.}} = 0.731$. 14. 1.
Page 199: $r = +\ 0.88$. 3. $r = -\ .89$.
Page 208: 2. (*a*) 158.
Page 210: 3. (*a*) 000004, (*b*) .04363.
4. $t = 5.9$.
Page 216: 3. 0.46.
Page 218: 1. $AF = 37.8 \pm .044$.
3. $AB = 11.44$.
Page 221: 3. $n > 156$.
Page 225: 3. $y = 0.1497 + 0.3782x + 0.9743x^2$, $\sigma_e = 2\ 07$, $r = 0.9997$.
Page 231: $\eta_{yx} = 0.732$; $\eta_{xy} = 0.700$, $r_{yx} = 0\ 682$. 2. $\eta_{yx} = 0.921$.
Page 261: 1. Er.Var. = 8 56.
Ex. Er. = 2.93. 3. B is better than A and C.
Page 269: 1. 2733.98 ± 13. 3. 2039.68 ± 106. 4. 13,932 ± 30 sq. ft. 10. $t = 21/19$. 11. .0197. 13. $t = 0.08/.037$. 18. $\sigma_{v1} = .4$, $\sigma_{v2} = .43$, $t < 1$, not sign. 19. No.
Page 278: 1. $t = 5.78$, sign.

ELEMENTARY MATHEMATICAL STATISTICS

CHAPTER 1

SUMMATIONS, CHARTS AND GRAPHS

In the study of statistics, symbols are used which enable one to shorten the writing of long expressions, certain series and summations. The following example will illustrate the meaning of a much-used symbol.

(1.1) The sum of the first nine positive integers

$$= 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9.$$

The left-hand member of the above equation may be designated by a symbol which will shorten the writing of this summation. This sum may be expressed by the symbol

$$\sum_{v=1}^{9} v,$$

where it is understood that v takes on integral values from 1 to 9 inclusive. This is read "the sum of the v's from 1 to 9 inclusive," or "the sum of the positive integers from 1 to 9 inclusive." The letter v is called the variable of summation and takes values as indicated. Equation (1.1) may be written as:

$$\sum_{v=1}^{9} v = 1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9,$$

or

$$\sum_{v=1}^{9} v = 45.$$

The symbol Σ is the Greek letter, capital sigma, and will be used to represent a sum or summation. It is sometimes read

"the sigma of the v's," or "the summation of the v's," or "the sum of the v's."

The equation $v = 1$ below the symbol Σ designates where the summation begins, and the number above sigma indicates where the summation stops. In the above example the sum begins with 1 and ends with 9.

The sum of the squares of the integers from 4 to 11 may be written as:

$$\sum_{v=4}^{11} v^2 = 4^2 + 5^2 + 6^2 + 7^2 + 8^2 + 9^2 + 10^2 + 11^2.$$

Numbers in the column below represent heights of 10 men

Heights of Men

v
$v_1 = 68.2$ in.
$v_2 = 67.4$
$v_3 = 69.3$
$v_4 = 72.0$
$v_5 = 71.2$
$v_6 = 66.9$
$v_7 = 68.8$
$v_8 = 69.0$
$v_9 = 70.2$
$v_{10} = 67.6$
$\Sigma v = 690.6$ in.

The sum of the heights in this column may be represented as a summation as follows:

$$\begin{aligned} \sum_{i=1}^{10} v_i &= v_1 + v_2 + v_3 + v_4 + v_5 + v_6 + v_7 + v_8 + v_9 + v_{10} \\ &= 68.2 \text{ in.} + 67.4 \text{ in.} + 69.3 \text{ in.} + \ldots + 70.2 \text{ in.} \\ &\quad + 67.6 \text{ in.} = 690.6 \text{ in.}, \end{aligned} \tag{1.2}$$

or

$$\sum_{i=1}^{10} v_i = 690.6 \text{ in.}$$

The first item in the column, 68.2 inches, is represented by v_1; the seventh item is represented by v_7, etc. In this case the

variable of summation is i; it runs from 1 to 10. The left-hand member of equation (1.2) is read "the sum of the v's from v_1 to v_{10} inclusive," or "the sum of the v's" or "the sum of the heights of the 10 men." Many times the numbers below and above sigma are omitted; the above summation may be written as:

$$\Sigma v = 690.6 \text{ in.}$$

Summations which follow also point out the use of Σ.

$$\sum_{k=2}^{8} \frac{1}{k^3} = \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{4^3} + \frac{1}{5^3} + \frac{1}{6^3} + \frac{1}{7^3} + \frac{1}{8^3};$$

$$\sum_{p=1}^{5} f(p) = f(1) + f(2) + f(3) + f(4) + f(5).$$

The variable of summation in the first sum is k; the variable of summation in the next is p.

To expand a sigma means to write out the summation in full; for example, $\sum_{u=2}^{7} u^4$ expanded is

$$\sum_{u=2}^{7} u^4 = 2^4 + 3^4 + 4^4 + 5^4 + 6^4 + 7^4.$$

Let it be required to write the following series by use of sigma:

$$4 \cdot 6 + 5 \cdot 7 + 6 \cdot 8 + 7 \cdot 9 + 8 \cdot 10 + 9 \cdot 11 + 10 \cdot 12.$$

The first numbers in the terms of this series begin with 4 and increase by unity up to and including 10. The second numbers in the terms begin with 6, increase from term to term by unity, and run to 12 inclusive. The series can be written as a sigma thus:

$$\sum_{x=4}^{10} x(x+2) = 4 \cdot 6 + 5 \cdot 7 + \ldots + 10 \cdot 12.$$

The following two theorems are indispensable.

THEOREM 1.1. The summation or the sigma of a constant times a function involving a variable of summation is equal to the constant times the sigma of the function, or

$$\sum_{i=1}^{n} a \cdot f(x_i) = a \cdot \sum_{i=1}^{n} f(x_i),$$

where a is a constant and $f(x)$ is a function of x.

Proof:

$$\sum_{i=1}^{n} af(x_i) = a\cdot f(x_1) + a\cdot f(x_2) + \ldots + a\cdot f(x_n)$$
$$= a\cdot[f(x_1) + f(x_2) + \ldots + f(x_n)] = a\sum_{i=1}^{n} f(x_i).$$

As an illustration

$$\sum_{i=1}^{n} 3(x_i^3 + x_i^2 - 2x_i) = 3\sum_{i=1}^{n} (x_i^3 + x_i^2 - 2x_i).$$

Theorem 1.2. The summation or the sigma of the sum of two functions involving a variable of summation is equal to the sum of the sigmas of the functions, or

$$\sum_{i=1}^{n} [f(x_i) + g(x_i)] = \sum_{i=1}^{n} f(x_i) + \sum_{i=1}^{n} g(x_i),$$

where $f(x)$ and $g(x)$ are functions of x.

Proof:

$$\sum_{i=1}^{n} [f(x_i) + g(x_i)] = f(x_1) + g(x_1) + f(x_2) + g(x_2) + \ldots$$
$$+ f(x_n) + g(x_n) = [f(x_1) + f(x_2) + \ldots + f(x_n)]$$
$$+ [g(x_1) + g(x_2) + \ldots + g(x_n)]$$
$$= \sum_{i=1}^{n} f(x_i) + \sum_{i=1}^{n} g(x_i).$$

PROBLEMS

Expand the following summations.

1. $\sum_{1}^{7} v^3.$
2. $\sum_{3}^{9} v(v+1).$
3. $\sum_{i=5}^{11} v_i f(v_i).$
4. $\sum_{x=11}^{16} \log(3x+4).$
5. $\sum_{r=0}^{8} \frac{1}{2^r}.$
6. $\sum_{d=-11}^{-2} d.$
7. $\sum_{t=51}^{62} p_t.$
8. $\sum_{v=-4}^{6} (v+3)f(v).$
9. $\sum_{i=17}^{24} v_i^2 f(v_i).$
10. $\sum_{s=3}^{6} 2^s.$
11. $\sum_{i=1}^{13} v_i^3 f(v_i).$
12. $\sum_{j=5}^{12} v_j w_j.$

Given that $\sum_{v=1}^{n} v = \frac{n}{2}(n+1)$; $\sum_{v=1}^{n} v^2 = \frac{n(n+1)(2n+1)}{6}$; $\sum_{v=1}^{n} v^3 = \frac{n^2(n+1)^2}{4}$, find the following:

13. $\sum_{v=1}^{n} 2v(v - \frac{1}{2})$.
14. $\sum_{v=1}^{n} (6v^2 - 2v)$.
15. $\sum_{v=1}^{n} v(v+1)(v+2)$.
16. $\sum_{v=1}^{n} (8v^3 - 12v^2 + 4v)$.
17. $\sum_{v=11}^{20} v^2$.

Write the following summations by use of sigma:

18. $7 \cdot 10 + 8 \cdot 11 + 9 \cdot 12 + 10 \cdot 13 + \ldots + 15 \cdot 18$.
19. $15f(15) + 16f(16) + \ldots + 42f(42)$.
20. $4^3 \cdot 5 + 5^3 \cdot 6 + \ldots + 12^3 \cdot 13$.
21. $\frac{1}{v_{10}} + \frac{1}{v_{11}} + \frac{1}{v_{12}} + \ldots + \frac{1}{v_{18}}$.
22. $v_6 f(v_6) + v_7 f(v_7) + \ldots + v_{30} f(v_{30})$.
23. $3 \cdot 7^3 + 5 \cdot 7^2 - 4 \cdot 7 + 3 \cdot 8^3 + 5 \cdot 8^2 - 4 \cdot 8 + 3 \cdot 9^3 + 5 \cdot 9^2 - 4 \cdot 9 + \ldots$
$\ldots + 3 \cdot 11^3 + 5 \cdot 11^2 - 4 \cdot 11$.

CHARTS AND GRAPHS

In analyzing statistical data it is often beneficial to exhibit results by means of graphs and charts, which enable one to grasp quickly the information contained in a mass of measurements, computations, and long columns of figures. Several types of charts and graphs which are used in many leading newspapers, scientific journals, and important reports will be explained in the remainder of this chapter.

Chart 1.1 contains 3 line charts pertaining to prices of wheat, corn, and oats since 1919. These broken lines allow the eye to see at once trends of prices of these essential commodities before, during, and after the depression. It is interesting to note that decreases in the price of one often accompanied decreases in the other two. Wheat prices were relatively higher in 1920 than prices of corn and oats. On examining these line diagrams it will be seen that prices for these grains started to drop as early as 1924. A marked increase began in 1933.

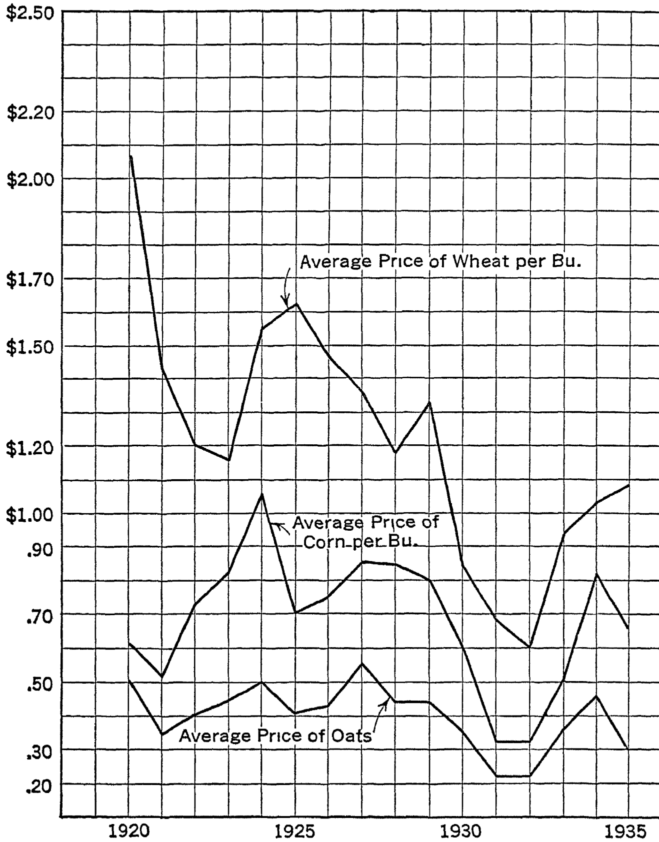


CHART 1.1.—Line charts showing average prices of wheat, corn and oats from 1920 to 1935. (*Agriculture Statistics of U. S. A.*, 1937.)

Bar diagrams often furnish adequate methods of presenting facts which enable one to make comparisons immediately. Charts 1.2 and 1.3 show production of cars and trucks and potatoes since 1928 and 1927 respectively. There was a drop in production of cars and trucks from 1929 to 1932 and an increase from 1932 to 1936. Production of potatoes (as chart 1.3 shows) did not have such great decreases and increases, as the heights of the bars show. The depression did not affect the number of bushels of potatoes grown as much as the number of cars and trucks manufactured. The peak for the automobile industry was in 1929; that for pota-

toes was in 1928. The space between bars should be about one-half the width of the bar.

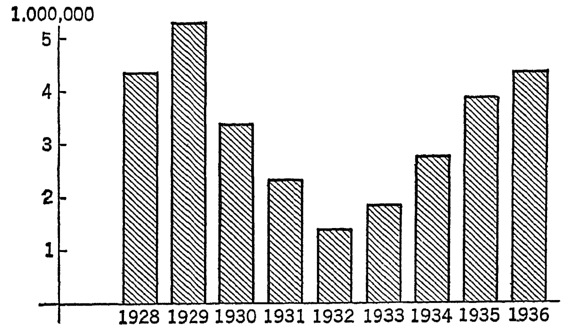


CHART 1.2.—Production of cars and trucks in United States and Canada. (*28th Annual Report of General Motors Corporation.*)

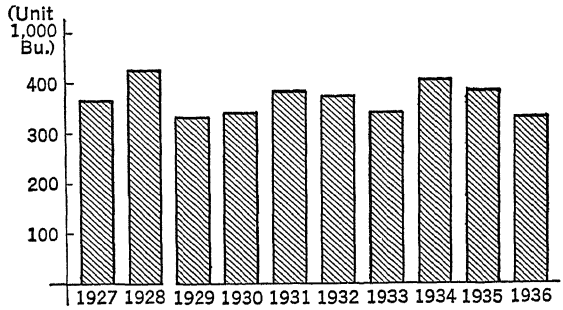


CHART 1.3.—Bar diagram showing average production of potatoes for U. S. A. from 1927 to 1936. (*Agriculture Statistics of U. S. A.*, 1937.)

Pie charts are useful for comparing percentages and the different parts of a whole. Chart 1.4 exhibits percentages of cotton production for 1935–36 and enables one to make rapid comparisons. Egypt and Brazil produced that year the same amount of cotton; China ranked third. The United States of America produced 44.3 per cent of all cotton grown in that period.

Chart 1.5 presents percentages of the divisions of the student body at one of the large agricultural colleges in America. This type of chart is easy to construct and can be interpreted readily.

Three-dimensional graphs can be used effectively to portray contrasts. Charts 1.6 and 1.7 are self-explanatory. Often pic-

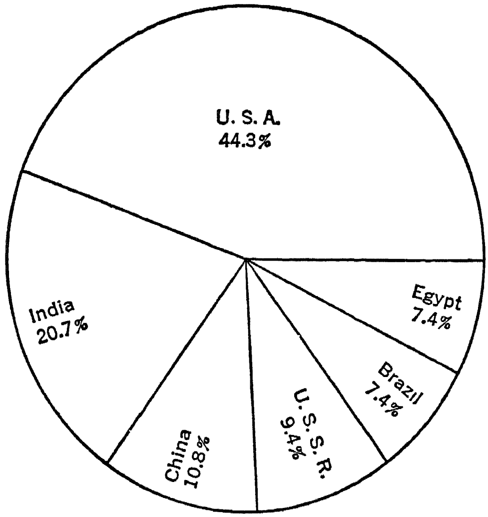


CHART 1.4 —Pie chart showing percentages of production of cotton during 1935–36.

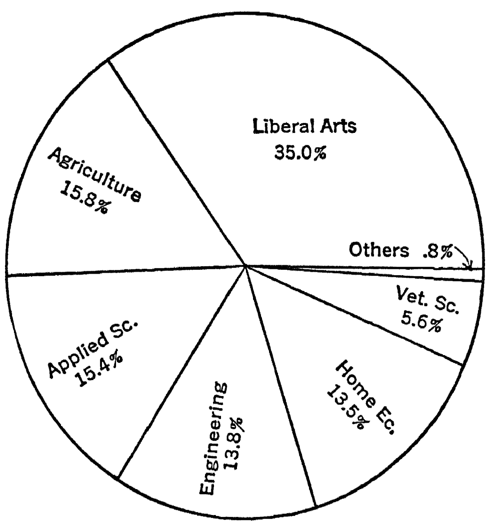


CHART 1.5.—Percentages of students in Michigan State College in 1935–36.

tures of objects will catch the eye where other types of charts will not. Many interesting bits of information have been given to the public through three-dimensional diagrams.

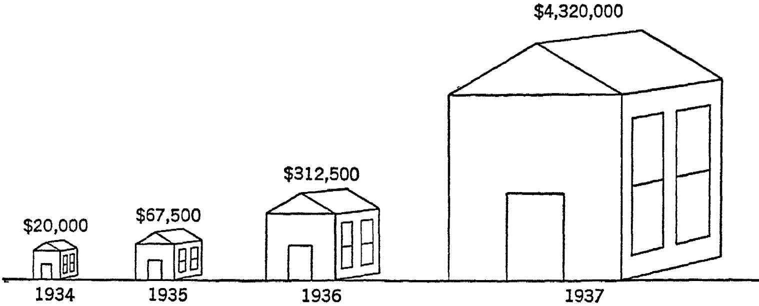


CHART 1.6.—Three-dimensional chart showing amount of building in a certain city.

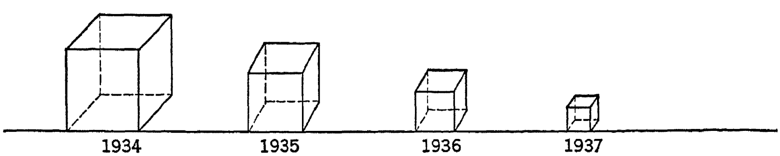


CHART 1.7.—Relative amounts of ice sold by all ice companies in a certain city from 1934 to 1937.

Zee charts set out in graphic form as a rule 3 phases of a business or project. This type is shown in chart 1.8; it contains monthly sales, cumulative sales, and 12-month moving total sales. The three lines look like the letter Z. The cumulative line is formed by plotting sales for January; January and February; January, February, and March; etc. The moving 12-month total sales line is formed by plotting the sales for the past 12 months. The managers can see three important phases of their business for the year. In this chart the moving 12-month total sales increased from December to June. The graph of cumulative sales is nearly a straight line. Monthly sales fell off considerably during July, August, and September.

Belt or strip charts can be employed when it is desirous to present several phases of a business or project. Chart 1.9 shows relief cases in Michigan for 1936 and part of 1937. The various belts or strips make vivid the relief program in Michigan. Aid to depen-

dent children and the blind started in October, as is seen in the chart. Sometimes different colors are used for the different strips.

There are many kinds of geographic charts. One is shown in chart 1.10, which indicates the percentages of the population of

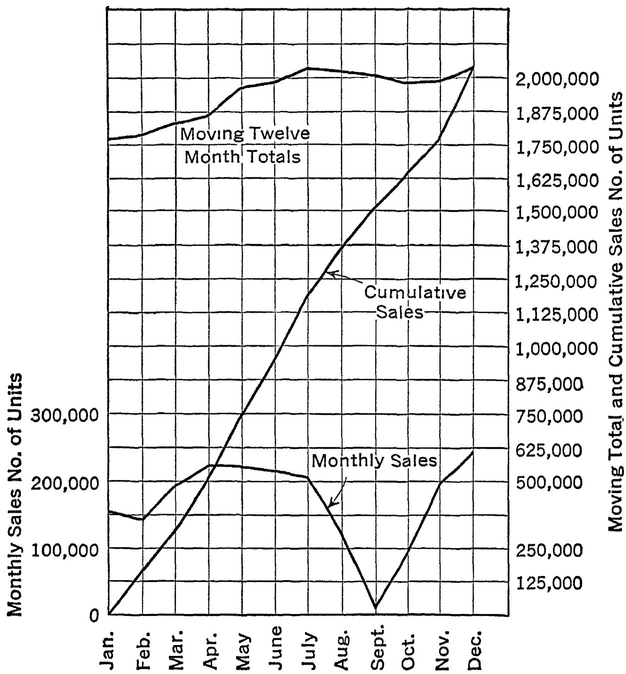


CHART 1.8.—Zee chart showing total unit sales of General Motors cars and trucks in 1936. (*28th Annual Report of General Motors Corporation.*)

Michigan by counties on relief. Well-constructed maps showing routes and all towns in which deliveries are made can be of much help to the superintendent of a business firm in assisting him to visualize his territory and the extent of his business. Sales for different districts can be shown on such maps. This type of chart is difficult to construct but when well drawn reveals a great deal of information.

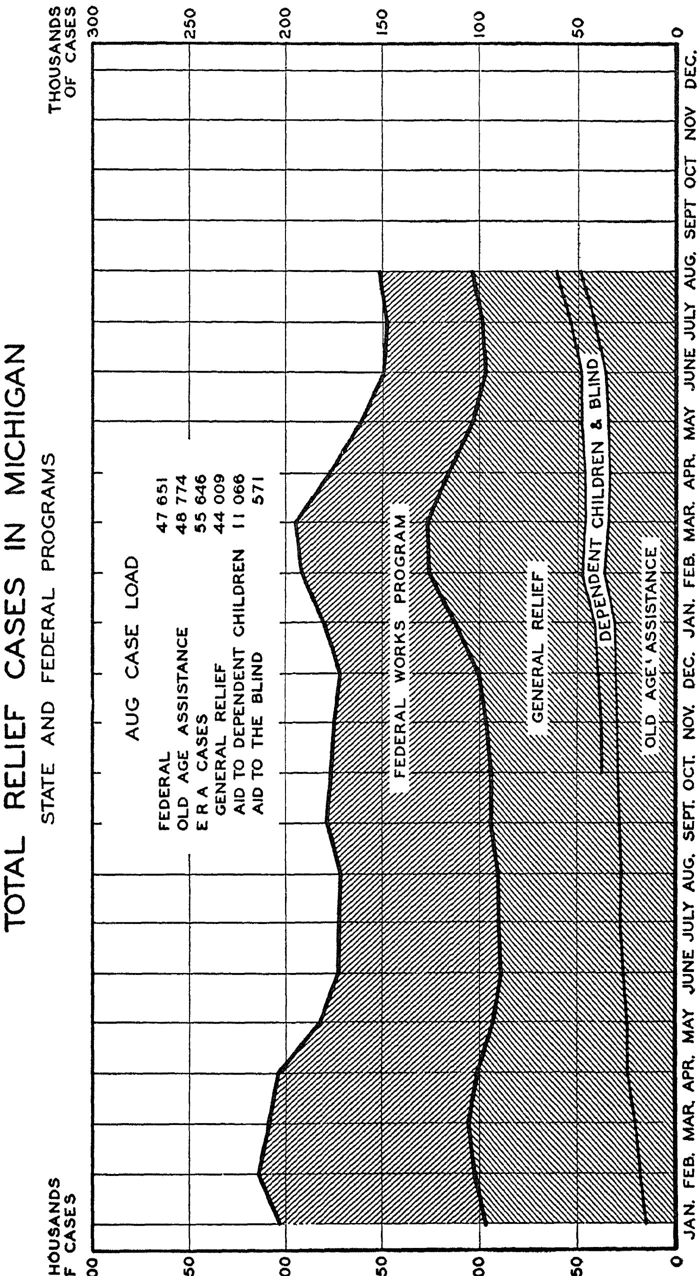


CHART 1.9.—Belt chart showing relief cases in Michigan. (*Monthly Bulletin on Public Relief Statistics*, Aug., 1937, of Michigan Emergency Relief Administration)

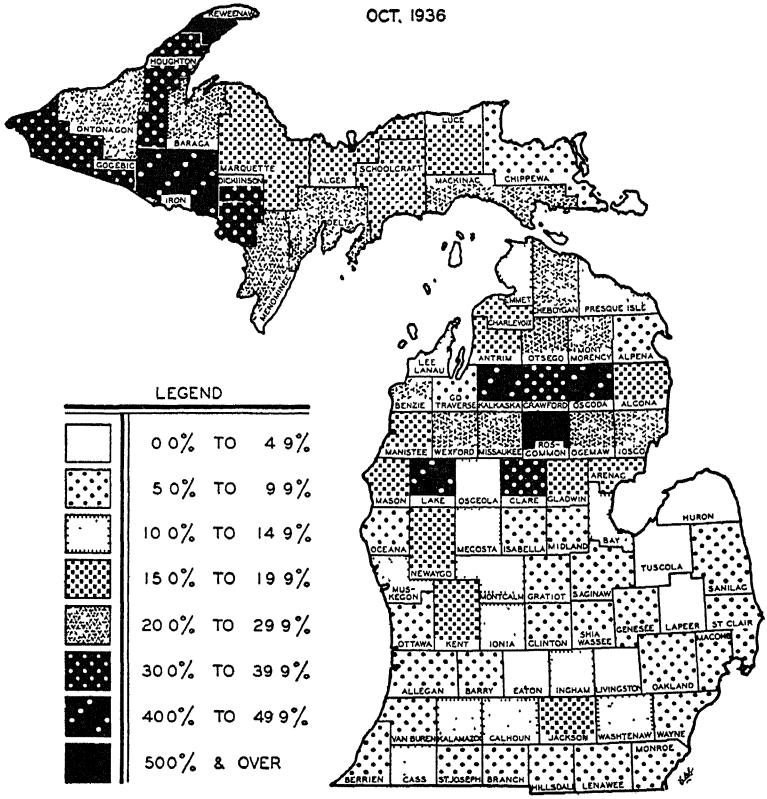


CHART 1.10.—Geographic chart showing per cent of population in Michigan on relief—cases on state and Federal relief programs. (*Monthly Bulletin on Public Relief Statistics*, Oct., 1936.)

PROBLEMS

1. Construct line charts for the following data:

	Number of Hogs on Farms, Millions	Number of Cattle and Calves on Farms, Millions
1920	60 2	70 4
1921	58.9	68 8
1922	59.8	68.8
1923	69.3	67 5
1924	66.6	66.0
1925	55.8	63 4
1926	52.1	60 6
1927	55.5	58.2
1928	61.9	57 3
1929	59.0	58.9
1930	55.7	61 0
1931	54.8	63.0
1932	59.3	65.8
1933	62.1	70 2
1934	58.6	74 3
1935	39.0	68.5
1936	42.8	68 2

This material was taken from the *Agriculture Statistics of U. S. A.*, 1937.

2. Construct line charts on one page for the following materials taken from the *Agriculture Statistics of U. S. A.*, 1937.

	Average Price of Eggs Per Dozen, Cents	Average Price of Butter per Pound, Cents	Average Price of Bread per Pound, Cents
1925	30.4	45.0	9.4
1926	28.8	44.4	9.4
1927	25.0	47.3	9.2
1928	28.0	47 4	9.1
1929	29.9	45.0	8.9
1930	23.7	36.5	8.2
1931	17.5	28 3	7.2
1932	14.2	21.0	6.6
1933	13.8	21.7	7.8
1934	17.0	25.7	8.3
1935	33.4	29.8	8.3
1936	21.7	33.1	8.2

3. Construct a bar chart for the following data:

Farm Wage Rates Per Month Without Board

Year	
1927	$48 63
1928	48.65
1929	49.08
1930	44.59
1931	35 03
1932	26 67
1933	24 51
1934	27 17
1935	29 48
1936	31 82

4. Compare the production of potatoes for Michigan and Maine by a bar chart.

	Michigan	Maine
Year	Millions of Bushels	
1927	23 1	37 4
1928	31 4	39 4
1929	15.9	49.9
1930	14 3	45 3
1931	23 8	50 3
1932	29 9	39 5
1933	20.7	42 0
1934	34 3	55.2
1935	28.1	38.4
1936	26.1	44.0

5. Construct pie charts for the following data:

Principal Exporting Countries of Cotton (1000 Bales) 1935-36

U. S. A.	British India	Egypt	Brazil	Argentina
6,397	2,999	1,688	650	202

6. Construct a pie chart for the following data pertaining to how a certain family spent its income. Salary $3,600.

Payments on home.	$900.00
Gas, light, water, fuel, telephone.	270.00
Taxes, upkeep on home, insurance.	540.00
Food.	600.00

Clothes	$150 00
Upkeep of car, trips	400 00
Installments on furniture	300 00
Church, charity	150.00
Pleasure, journals, books	200.00
Miscellaneous	90.00

7. Construct a Zee chart for the following data:

MICHIGAN EMERGENCY RELIEF STATISTICS
NUMBER OF PEOPLE ON RELIEF *

	1935	1936
Jan.	237,231	82,764
Feb.	216,634	88,862
Mar.	200,984	88,753
Apr.	190,600	81,075
May	181,295	71,195
June	183,468	64,146
July	183,732	61,451
Aug.	180,725	62,922
Sep.	175,284	65,191
Oct.	163,867	64,336
Nov.	142,635	65,931
Dec.	107,494	70,277

* This material was taken from *Monthly Bulletin on Public Relief Statistics* published by Michigan Emergency Relief Administration, Vol IV, No. 9, 1937.

8. Construct a belt or strip chart for the following sales for a book store:

	NO. BIBLES	FICTION	NO. TEXTBOOKS	ALL OTHERS
Jan.	37	20	55	61
Feb.	23	16	31	38
Mar.	28	30	48	56
Apr.	35	27	36	70
May	30	22	16	53
June	23	15	62	39
July	20	11	19	32
Aug.	14	9	8	26
Sep.	18	24	91	72
Oct.	29	38	63	81
Nov.	48	36	31	90
Dec.	63	50	13	134

9. Bring to class two geographic charts which you have seen in the local newspapers or current journals.

LOGARITHMIC GRAPHS

Sometimes it becomes necessary to employ logarithmic paper for presenting data in graphic form. For semilogarithmic paper one of the axes is divided into equal units while the other axis is

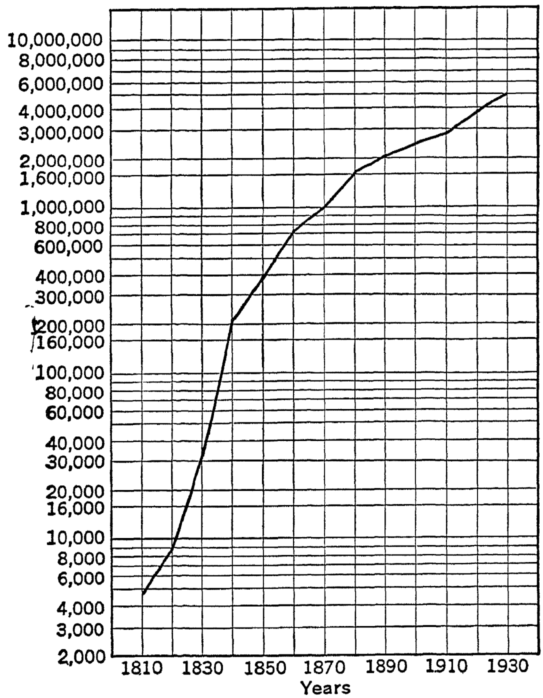


CHART 1.11.—Population of Michigan from 1810 to 1930 plotted on semilogarithmic paper.

divided into logarithm units. The horizontal axis in chart 1.11 represents years; the vertical axis represents the logarithms of the population figures for Michigan from 1810 to 1930. The graph is nearly a straight line, showing that the rate of increase is nearly a constant for decades. It is difficult to place on the same page census figures for Michigan in 1810 and those for the last two decades, when plotted on an arithmetic scale.

Chart 1.12 shows the compound amount on $10 compounded

annually at 10 per cent plotted on semilogarithmic paper. The line is straight when plotted on this kind of paper.

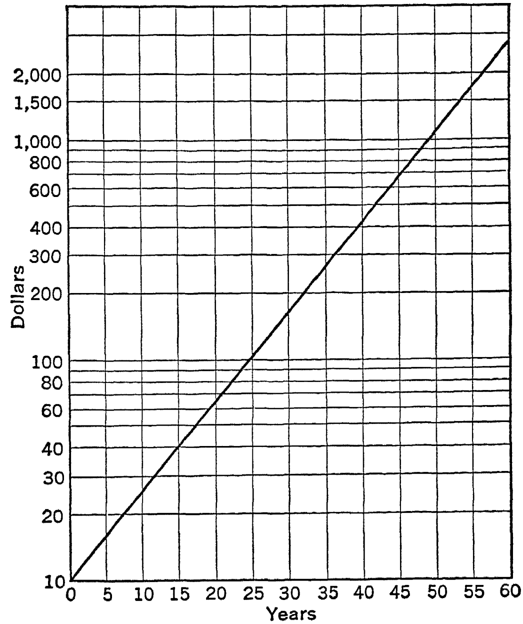


CHART 1.12 —Compound amount on \$10.00 compounded annually at 10 per cent plotted on semilogarithmic paper.

Relations similar to the following can be represented by straight lines if drawn on semilogarithmic paper.

$$y = a \cdot b^x,$$

or

$$\log y = \log a + x \log b,$$

where a and b are positive constants. The constant a may be equal to 1. The constant a is equal to y when x is equal to zero. The compound amount law, represented in chart 1.12, is

$$y = 10(1.10)^x,$$

or

$$\log y = \log 10 + x \log (1.10);$$

here $a = 10$, $b = 1.10$, $y =$ the compound amount and $x =$ the length of time in years.

Double logarithmic paper can be used to plot relations similar to the following

$$y = a \cdot x^b,$$

or

$$\log y = \log a + b \log x,$$

where a and b are constants. The constant a must be positive; it is equal to y when $x = 1$.

Chart 1.13 shows the speed of trackmen for various distances plotted on double logarithmic paper. The line is nearly straight with a slight break at 880 yards.

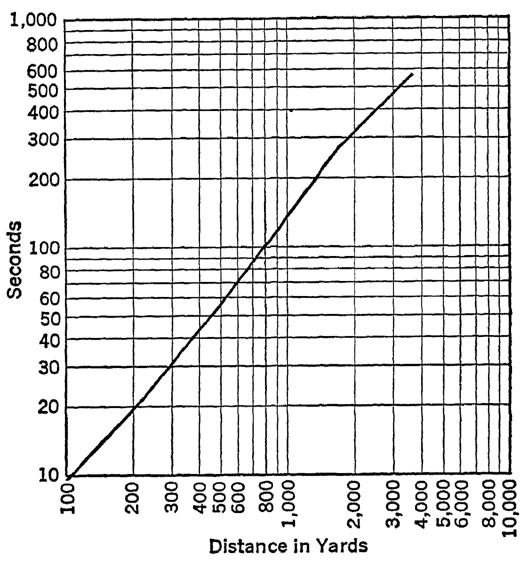


CHART 1.13.—Speed of trackmen at Michigan State College for various distances plotted on double logarithmic paper.

PROBLEMS*

1. Plot on semilogarithmic paper the following census figures for the U. S. A. Try to predict what the population will be in 1950.

Year	Population
1790	3,929,000
1800	5,308,000
1810	7,240,000
1820	19,638,000
1830	12,866,000
1840	17,069,000
1850	23,192,000
1860	31,443,000
1870	38,558,000
1880	50,156,000
1890	62,948,000
1900	75,995,000
1910	91,972,000
1920	105,711,000
1930	122,775,000

2. Plot on single logarithmic paper the following data:

Angle of Contact between Belt and Pulley from Pull

Angle, radians	1	2	3	4	5
P, pounds	4.5	6.5	9.5	14	20

3. Plot on double logarithmic paper the following data:

Track and Field Records of Olympic Men

Distance, meters	100	200	400
Time	10.3 sec.	21 2 sec.	46.2 sec.
Distance, meters	800	1,500	5,000
Time	1 min. 49.8 sec.	3 min. 51 2 sec.	14 min. 30 sec.

4. Plot the following data on double logarithmic paper:

Age of tree, years	5	10	25	50	75	100
Height, feet	6.5	11	22	38	50	64

* Problems 1, 2, 3, and 4 were taken by permission from "Statistical Data" by S. E. Crowe.

CHAPTER 2

STATISTICAL AVERAGES

ARITHMETIC MEAN

Since averages play important roles in statistical work, it is essential to be familiar with some of the more common ones and to know how to obtain them by the shortest methods. Five averages are introduced in this chapter together with methods for securing them. The first to be considered is the arithmetic average or the arithmetic mean.

The following column contains the number of children in families living in a certain section.

No. of Children per Family, v
$v_1 = 1$
$v_2 = 3$
$v_3 = 5$
$v_4 = 8$
$v_5 = 4$
$v_6 = 2$
$v_7 = 6$
$v_8 = 5$
$v_9 = 4$
$v_{10} = 9$
$v_{11} = 11$
$v_{12} = 6$
$v_{13} = 4$
$v_{14} = 5$
$v_{15} = 2$
$\Sigma v = 75$

The average number of children per family for this section is equal to

$$\text{Mean} = M_v = \frac{1 + 3 + 5 + \ldots + 5 + 2}{15}$$
$$= \frac{\Sigma v}{15} = 5 \text{ children.}$$

Five is the average number of children per family and is a good representative of the number of children on the average for each family for this group of families. The mean of a set of items in many instances is a good representative of the items and shows a central tendency of the items.

Let the quantity v_i represent the ith item or variate in a set of variates. In the above example $v_1 = 1$, $v_2 = 3$, $v_3 = 5$, etc. The quantity v is a variable, and the particular values of v are called variates. In the above illustration the number of children per family is a variable; the actual values observed in the families are variates.

The arithmetic mean of a set of variates is equal to the sum of the variates divided by the number of variates, or

$$\text{(2.1)} \qquad \text{Mean} = M_v = \frac{v_1 + v_2 + \ldots + v_n}{n} = \frac{\Sigma v}{n},$$

where n is the number of variates in the set.

PROBLEMS

1. Find the mean of the following figures, which are the number of seeds in pods of a certain plant. 7, 11, 14, 9, 10, 12, 8, 11, 10, 13, 12, 12, 14, 10, 11, 9, 13, 11, 12, 11.

2. The next table exhibits data pertaining to rainfall at Grand Rapids and Marquette, Michigan, from 1912 to 1935 during December. (Taken from the *Year Book of the U.S.A. Dept. of Agri.*, 1935.)

	Precipitation in Inches	
Year	Grand Rapids	Marquette
1912	1 32	2.42
1913	0 31	0 94
1914	1.89	0.85
1915	1.22	2.17
1916	3.81	3.09
1917	0 82	3.86
1918	4.02	2.94
1919	1.19	1.89
1920	4 19	2 27
1921	4 14	2 04
1922	1.40	1.14
1923	2.18	1.52
1924	1.66	1.53
1925	1.52	2 80
1926	1.86	2 36
1927	3.16	3.19
1928	2.79	1.13
1929	2.71	3.47
1930	1 34	1 97
1931	2.64	1.09
1932	2.68	1.60
1933	1 74	2.58
1934	1 51	2.30
1935	1.65	2.91

Find the average amount of rainfall during December for the two cities, and compare results. Locate the cities on a map, and discuss results.

3. Gallons of gasoline sold by a certain filling station during June: 480, 459, 468, 477, 453, 490, 450, 441, 484, 472, 460, 449, 502, 465, 468, 481, 430, 463, 498, 478, 405, 446, 473, 453, 469, 438, 458, 467, 420, 439. Find the average output of gasoline per day during June for this station. Plot these data on a vertical scale, draw a line representing the mean, and find the number of items below and above this line.

4. Number of eggs laid by 40 hens during 20 weeks.

Weeks	No. of Eggs	Weeks	No. of Eggs
1	180	11	210
2	176	12	240
3	168	13	209
4	164	14	256
5	180	15	236
6	177	16	227
7	193	17	240
8	196	18	220
9	212	19	213
10	200	20	190

Find the average number of eggs laid by the hens per week and the average number laid per hen per week.

ARITHMETIC AVERAGE FOR FREQUENCY DISTRIBUTIONS

Data in problem 1 of the last section can be arranged into a frequency distribution by collecting the items. This is how the first two columns in the next table arose.

	No. of Seeds per Pod v	No. of Pods or Frequency $f(v)$	$v \cdot f(v)$
	7	1	7
	8	1	8
	9	2	18
	10	3	30
	11	5	55
	12	4	48
	13	2	26
	14	2	28
Total		$\Sigma f(v) = 20$	$\Sigma v \cdot f(v) = 220$

This distribution of the items is called a frequency distribution because the items are listed with their respective frequencies. Number 11 appears 5 times, 12 appears 4 times, etc. Frequencies are designated by $f(v)$ and are found in that column. The column headed by $v \cdot f(v)$ contains products of the variates by their respective frequencies. For example, 12 occurred 4 times, hence the sum of these four 12's is 48; this 48 appears in the $v \cdot f(v)$ column. The sum of the numbers in the $v \cdot f(v)$ column is the sum of all items

and is the same as was obtained in the first problem of the last section. The mean of the items can be written in terms of frequencies as follows:

$$\text{Mean} = M_v = \frac{\text{Sum of the items}}{\text{No. of items}} = \frac{\Sigma v \cdot f(v)}{\Sigma f(v)}.$$

The total number of items, n, is equal to the sum of the frequencies, or

$$n = \Sigma f(v_i) = \Sigma f(v).$$

The mean for the above example is

$$M_v = \frac{\Sigma v \cdot f(v)}{\Sigma f(v)} = \frac{220}{20} = 11 \text{ seeds.}$$

A frequency distribution enables one to put a great deal in a little space and aids one in calculating the average if several of the variates are repeated.

PROBLEMS

1. The following data give the number of ligulate flowers or ray flowers on heads of oxeye daisies growing on a certain plot: 12, 16, 15, 20, 18, 17, 18, 21, 18, 15, 17, 19, 20, 14, 18, 16, 15, 13, 19, 18, 17, 14, 21, 17, 16, 14, 18, 19, 20, 13, 15, 16, 17, 19, 16, 18, 19, 17, 16, 19, 18, 18, 18, 17, 14, 16, 18, 17, 16, 18, 19, 15, 20, 19, 18, 17, 16, 15, 18, 19, 19, 17, 16, 16, 15, 18, 19, 20, 18, 20, 15, 16, 19, 18, 18, 17, 17, 17, 16, 19, 16, 18, 17, 18, 21, 22, 20, 19, 16, 18, 15, 21, 14, 19, 19, 19, 17, 18, 16, 17, 16, 17, 15. Arrange the items in a frequency distribution, and find the average number of ligulate flowers per daisy.

2. Below is a frequency distribution of the number of loaves of bread sold by a grocery store for 92 days.

No. of Loaves Sold per Day v	No. of Days $f(v)$	No. of Loaves Sold per Day v	No. of Days $f(v)$
175	1	219	5
182	4	222	3
186	9	228	3
193	17	231	2
197	22	235	1
201	14	237	1
204	10		

Find the average number of loaves sold per day.

3. The next table contains scores made by seventh-grade students on a certain test. Find the average score.

Scores v	No. of Students $f(v)$	Scores v	No. of Students $f(v)$	Scores v	No. of Students $f(v)$
49	2	65	42	81	68
50	3	66	45	82	60
51	3	67	52	83	53
52	2	68	59	84	46
53	3	69	65	85	38
54	5	70	66	86	25
55	8	71	72	87	17
56	10	72	75	88	8
57	14	73	79	89	3
58	16	74	83	90	4
59	19	75	90	91	3
60	23	76	88	92	2
61	25	77	91	93	3
62	28	78	90	94	2
63	32	79	84	95	1
64	37	80	75	96	1

If A is considered to be from 90 to 100, B from 80 to 90, C from 70 to 80, D from 60 to 70, and E any score below 60, find the percentages of A's, B's, C's, D's, and E's.

INDIRECT METHOD OF COMPUTING THE MEAN

Figures in the first column of the next table are the number of gallons of gasoline sold each day for 10 days.

No. of Gallons of Gas Sold per Day v	d $v-460$	d $v-455$
450	-10	-5
445	-15	-10
462	2	7
470	10	15
449	-11	-6
455	-5	0
458	-2	3
460	0	5
464	4	9
447	-13	-8
$\Sigma v = 4{,}560$	$\Sigma d = -40$	$\Sigma d = 10$

From each item in the first column subtract a provisional or guessed mean, 460, and place these differences in the first d column. This column gives deviations of the variates from this provisional mean The average of the deviations of the variates from 460 is -4 gallons, hence the average number of gallons sold per day is

$$M_v = 460 + \frac{\Sigma d}{10} = 460 - 4 = 456 \text{ gallons.}$$

Numbers in the third column are deviations of the variates from another provisional mean, 455 gallons; the average number of gallons sold per day is

$$M_v = 455 + \frac{\Sigma d}{10} = 455 + 1 = 456 \text{ gallons,}$$

as was obtained when 460 gallons was used as provisional mean.

The object of this method of finding the mean is to reduce the size of the variates and hence to reduce the amount of computing. Numbers in the second and third columns are smaller than those in the first. If the provisional mean is chosen as a number near the middle of the range of the items, in most cases, deviations of the items from this provisional mean will be small compared to the original items. The next theorem shows that this method holds for any provisional mean.

THEOREM 2.1. The mean of a set of variates is equal to the provisional mean, h, plus the average of deviations of the variates from the provisional mean, or

$$M_v = h + M_d. \tag{2.2}$$

PROOF: By definition, deviations of the variates from the provisional mean, h, are

$$\begin{aligned} d_1 &= v_1 - h \\ d_2 &= v_2 - h \\ d_3 &= v_3 - h \\ \cdot\cdot\; &\;\;\cdot\cdot\cdot\cdot\cdot\cdot \\ \cdot\cdot\; &\;\;\cdot\cdot\cdot\cdot\cdot\cdot \\ \cdot\cdot\; &\;\;\cdot\cdot\cdot\cdot\cdot\cdot \\ d_n &= v_n - h \\ \hline \text{Total}\quad \Sigma d &= \Sigma v - n \cdot h \end{aligned}$$

Divide the above totals by n, the number of variates; this becomes

$$\frac{\Sigma d}{n} = \frac{\Sigma v}{n} - \frac{n \cdot h}{n}, \quad \text{or} \quad M_d = M_v - h,$$

or

$$M_v = h + M_d.$$

PROBLEM

1. The following table contains attendance at a Sunday school for one quarter. By use of a provisional mean find the average attendance.

No. of Students per Sunday	No. of Students per Sunday
137	158
130	149
146	151
129	158
143	155
146	147
149	

The indirect method of finding the mean can be used in finding the mean of a frequency distribution, as the next illustration will show. The table below gives the number of eggs laid by 267 pheasant hens.

No. of Eggs v	No. of Pheasants $f(v)$	d $v-13$	$d \cdot f(d)$
7	1	−6	−6
8	7	−5	−35
9	13	−4	−52
10	25	−3	−75
11	34	−2	−68
12	59	−1	−59
13	55	0	−295
14	41	1	41
15	23	2	46
16	6	3	18
17	2	4	8
18	1	5	5
	267		+118
			−177

Let the provisional mean, h, be equal to 13 eggs; deviations of the variates from this provisional mean are listed in the third column. Products of these deviations by their respective frequencies are listed in the last column. The mean, according to formula (2.2), is

$$M_v = h + M_d = 13 - \frac{177}{267} = 13 - 0.66 = 12.34 \text{ eggs.}$$

Theorem 2.1 is one of the most important theorems in statistics.

PROBLEMS

1. By use of a provisional mean compute the mean of the following distribution of petals on the ordinary buttercup (*Ranunculus bulbosus*).

No. of Petals v	No. of Flowers $f(v)$	No. of Petals v	No. of Flowers $f(v)$
5	40	11	177
6	52	12	104
7	126	13	35
8	165	14	8
9	204	15	4
10	215		1,130

2. The table below contains the number of strokes used by a golfer for 100 days on a certain course. Find by use of a provisional mean his average score.

No. of Strokes for 18 Holes	No. of Days	No. of Strokes for 18 Holes	No. of Days
80	1	91	12
83	3	92	8
85	10	93	5
86	17	94	3
88	26	95	1
89	14		

3. If the mean of 179 weights is 138.3 pounds, what is the sum of the weights?

4. If the sum of lung capacities of pupils in a class is 5,625 cubic inches and the mean lung capacity for these pupils is 225 cubic inches, find the number of pupils in the class.

5. If the mean of 75 items is 40.4 quarts and the mean of 25 items is 39.1 quarts, find the mean of the 100 items.

6. The table below contains the distribution of stigmatic rays on seed capsules of Shirley poppies. (See *Biometrika,* II, 1902, p. 89.)

No. of Rays	No. of Capsules	No. of Rays	No. of Capsules
6	3	14	302
7	11	15	234
8	38	16	128
9	106	17	50
10	152	18	19
11	238	19	3
12	305	20	1
13	315		1,905

Find the average number of stigmatic rays per poppy seed capsule by use of a provisional mean.

THE MEAN FOR THE COMBINATION OF SETS

THEOREM 2.2. If the mean of n_1 items is M_1 and the mean of n_2 items is M_2, then the mean of the combined set of $n_1 + n_2$ items is

$$M_{1+2} = \frac{n_1M_1 + n_2M_2}{n_1 + n_2}. \tag{2.3}$$

PROOF: Since the numerator in the above equation is the sum of the items in both sets and the denominator is the total number of items, the mean of the combined sets is equal to the right-hand member of the equation.

In general, if the mean of n_1 items is M_1, the mean of n_2 items is M_2, the mean of n_3 items is M_3, . . . , and the mean of n_s items is M_s, then the mean of the items in all the sets is

$$M_{1+2+\ldots+s} = \frac{n_1M_1 + n_2M_2 + n_3M_3 + \ldots + n_sM_s}{n_1 + n_2 + n_3 + \ldots n_s}. \tag{2.4}$$

PROBLEMS

1. The mean score on an arithmetic test of

55 students in Ward 1 is 76.2,
69 " " " 2 " 74.3,
30 " " " 3 " 78.1,
44 " " " 4 " 75.8.

Find the average score on this test for the city.

2. The average weight of the 35 men in dormitory A is 140.4 pounds,
" " " " " 67 " " " B " 139.3 " ,
" " " " " 96 " " " C " 138.6 " .

Dormitory C contains 22 Japanese men whose average weight is 136.3 pounds. Find the average weight of the men in all dormitories without the Japanese.

3. Average amount of gas sold per day by Station 1 was 430 gallons,
" " " " " " " " " 2 " 362 " ,
" " " " " " " " " 3 " 284 " ,
" " " " " " " " " 4 " 685 " ,
" " " " " " " " " 5 " 598 " ,
" " " " " " " " " 6 " 507 " ,
" " " " " " " " " 7 " 706 " ,
" " " " " " " " " 8 " 843 " ,
" " " " " " " " " 9 " 389 " .

If these data were for the month of May, find the average number of gallons of gas sold by the company owning these stations per day. Find the amount sold during the month.

4. How should formula (2.3) be written if n_2 is a subset of n_1, and it is required to find the mean of the $n_1 - n_2$ items?

5. Express n in terms of the sum of the variates and the mean.

6. The following data relate to sales of stamps in a post office for one day.

Denomination of Stamps	No. Sold	Denomination of Stamps	No. Sold
$\frac{1}{2}$ cent	648	6 cents	89
1 "	1409	10 "	94
2 cents	1278	16 "	59
3 "	2941	25 "	37
5 "	703		

Find the amount received from stamps.

THE MEAN OF GROUPED DATA

Many times, labor is saved in computing when variates are divided into groups or classes. This will be illustrated by an example. The table below contains the distribution of weights of 100 male students. The original measurements were made to the nearest half pound. These weights have been grouped into classes of 10 pounds.

Weight of Students v	Class Marks x	Frequency $f(v)$	$x-134.75$ d	$d \cdot f(d)$
99 75–109 75	104 75	1	−3	− 3
109 75–119.75	114 75	9	−2	−18
119.75–129.75	124 75	23	−1	−23
129.75–139.75	134.75	25	0	−44
139–75–149.75	144.75	19	1	19
149 75–159 75	154.75	15	2	30
159 75–169.75	164.75	4	3	12
169.75–179.75	174.75	4	4	16
		100		+77
				+33

The lower closed limit of the second class is 109.75 pounds, because every weight which was less than this and greater than 99.75 pounds was put into the first class and all above this and less than 119.75 pounds was put into the second class. The person who weighed the students wanted weights to the nearest half-pound. If the scales showed a weight to be 109.83 pounds it was recorded as 110 pounds, as this was the weight to the nearest half pound; had the weight been 109.65 pounds it would have been recorded as 109.5 pounds. Any weight between 109.75 pounds and 110.25 pounds was recorded as 110 pounds; any weight between 137.25 pounds and 137.75 pounds was recorded as 137.5 pounds. After the weights were recorded they were grouped into classes of 10 pounds.

The class mark of a class is the average of the upper and lower closed limits of the class. The class mark of the third class is $\frac{119.75 + 129.75}{2} = 124.75$ pounds. Class marks are written in the second column. Class marks are representatives of the items in the respective classes. Weight, 124.75 pounds, represents the weight of each item in the third class. Some weights in the third class were less than this class mark, and some were greater. In grouping variates into classes it is assumed that the class mark is the average of the variates in the class. Of course, the class mark is the average of the variates in the class if there are the same number in the class above the class mark as below it, and equally

spaced. This happens often when there are large numbers of items in the classes and the classes have been made small. When variates have been grouped into classes the individual variates are not known; the class mark is the representative of all variates in the class. There are 9 weights in the second class; the original measurements of these are not known now. The class mark, 114.75 pounds, now represents these 9 measurements; this is the same as saying that each of the 9 measurements is 114.75 pounds. The distribution is now made up of class marks and their frequencies, which are the numbers of measurements in the respective classes. The frequency of class mark 114.75 pounds is 9 because there are 9 measurements in this class.

The mean of the distribution made up of class marks and their frequencies can be found by applying formula (2.2). Choose the class mark of the fourth class as provisional mean and place the deviations of class marks from this provisional mean in column four. Deviations, d, are now in terms of class units; that is, the difference between two consecutive class marks is 10 pounds or 1 class unit. The mean of the distribution made up of the d's is

$$M_d = \frac{\Sigma d \cdot f}{\Sigma f} = \frac{+33}{100} = 0.33 \text{ class unit.}$$

This 0.33 class unit is converted into pounds by multiplying by 10, the size of the class; hence the mean of the distribution made up of class marks is

$$(2.5) \qquad M_x = h + M_d \cdot w = 134.75 + (0.33)10 = 138.05 \text{ lb.},$$

where w is the width of the class. The mean of the x's, 138.05 pounds, is considered to be the mean of the grouped data, or the mean of the v's. It is not always the same as would be obtained by taking the average of the measurements before they were thrown into classes. Work in more advanced statistics shows that the mean of class marks and the mean of original measurements before they are grouped differ by a negligible quantity, provided the classes are not too large. Quantity w is the width of the class; it is 10 pounds in this case. Formula (2.5) is the same as formula (2.2) with $w = 1$. The object of grouping is to reduce the amount of labor in computing.

The degree of accuracy of measurements is very important, for this enables one to obtain the class limits. Suppose that measurements in the above illustration had been made to the nearest pound; then the classes would have been written as follows:

99.5–109.5
109.5–119.5
119.5–129.5
etc.

Had measurements been made to the nearest 0.1 pound the classes would have been written as:

99.95–109.95
109.95–119.95
etc.

It is very important to get the class limits correct, for they determine class marks which determine the mean of the distribution. One cannot measure anything to the nearest $\frac{1}{1,000,000,000,000}$ of a unit or to the nearest infinitesimal of a unit; hence the degree of accuracy of measurements must be given so that anyone may be able to group the data. An incorrect way of writing classes in the above example where measurements were made to the nearest 0.1 pound is as follows:

100–110
110–120
etc.

Classes in this case may be written in open limits as follows:

100–109.9
110–119.9
120–129.9
etc.

To find the closed limits from the open limits one decreases the lower open limit of the class by one-half of the degree of accuracy of the measurements and adds one-half of the degree of accuracy to the upper open limit. The lower closed limit of a class is the same as the upper closed limit of the preceding class; this is not true for open limits.

The following figures represent weights of 100 men. The original measurements were made to the nearest pound.

WEIGHTS	FREQUENCY	WEIGHTS	FREQUENCY	WEIGHTS	FREQUENCY	WEIGHTS	FREQUENCY
111	1	126	3	139	1	154	1
112	1	127	2	140	5	155	3
113	2	128	1	141	2	156	1
114	1	129	1	143	6	158	1
115	1	130	4	144	3	159	1
117	2	131	2	145	1	160	1
118	2	132	1	146	2	162	1
119	1	133	2	147	2	163	1
120	1	134	3	148	2	164	1
121	3	135	1	149	1	168	1
122	4	136	4	150	5	169	1
124	1	137	1	152	1	175	1
125	2	138	5	153	2	179	1

The above data were grouped in 2's, 3's, and so on, up through 10's. The lower limit of the first class was always 109.5 pounds. The means for the different ways of grouping are:

$$
\begin{array}{ll}
M_1 = 138.16 \text{ lb.}, & M_6 = 138.12 \text{ lb.}, \\
M_2 = 138.28 \text{ ''}, & M_7 = 138.42 \text{ ''}, \\
M_3 = 138.24 \text{ ''}, & M_8 = 138.38 \text{ ''}, \\
M_4 = 138.34 \text{ ''}, & M_9 = 138.12 \text{ ''}, \\
M_5 = 138.25 \text{ ''}, & M_{10} = 138.50 \text{ ''},
\end{array}
$$

where M_6 represents the mean when the size of the class was 6 pounds, etc. The mean of the ungrouped data is $M_1 = 138.16$ pounds, and is the true mean, provided the original measurements are considered exact. Means for grouping in 6's and 9's are nearer the "true" mean than means for other groupings.

This example shows that means obtained from grouped data give results which differ very little from the true mean; in this case they differ by less than 0.35 pound from the "real" mean. Had there been more items in the distribution there would have been better results. This example shows that in grouping one does not always secure the true average but an average which is approximately equal to the true average. Of course, the true mean may never be obtained because the measurements were made only to the nearest pound.

PROBLEMS

1. Group the following freshmen mathematics averages in groups of 5, and find the mean of the distribution. Use 49.5 as the lower limit of the first class: 62, 82, 78, 78, 82, 78, 92, 68, 85, 75, 60, 85, 50, 92, 72, 88, 72, 70, 75, 82, 92, 60, 72, 95, 95, 95, 75, 78, 75, 95, 83, 85, 75, 70, 68, 85, 82, 65, 88, 78, 82, 92, 95, 92, 82, 78, 82, 82, 74, 75, 82, 55, 68, 92, 68, 92, 72, 75, 78, 62, 95, 72, 83, 83, 65, 75, 82, 88, 72, 78, 78, 95, 77.

2. The following table gives the frequency distribution of rays per inflorescence of *Zizia aurea* (Golden Alexander). Find the mean number of rays per cluster.

No. of Rays per Cluster	Frequency	No. of Rays per Cluster	Frequency
10–11	21	20–21	65
12–13	39	22–23	22
14–15	98	24–25	6
16–17	127	26–27	11
18–19	111		500

3. The following table contains lung capacities, to the nearest cubic inch, of 536 women students at the University of Michigan. Find the mean, and plot the distribution.

Lung Capacity	No. of Women	Lung Capacity	No. of Women
79.5– 99 5	2	179.5–199.5	114
99 5–119 5	14	199.5–219 5	53
119.5–139 5	78	219.5–239.5	30
139 5–159 5	78	239.5–259 5	9
159.5–179 5	158		536

4. The table below gives weights of 194 women students at the University of Michigan, measurements being made to the nearest 0.1 pound. Find the mean.

Weights	Frequency	Weights	Frequency
79.45– 89.45	2	139.45–149.45	11
89.45– 99.45	18	149.45–159.45	6
99.45–109.45	28	159.45–169.45	3
109.45–119.45	55	169.45–179.45	2
119.45–129.45	49	179.45–189.45	1
129.45–139.45	19		194

5. The following is a distribution of pedicels per inflorescence of *Daucus carota* (wild carrot) from Michigan. Find the mean. Make the classes twice as large and find the mean. Compare this with the mean found before the classes were doubled in size. Which is nearer the true mean?

No. of Pedicels per Cluster	No. of Clusters	No. of Pedicels per Cluster	No. of Clusters
25–29	1	60– 64	119
30–34	19	65– 69	74
35–39	44	70– 74	34
40–44	133	75– 79	13
45–49	152	80– 84	15
50–54	230	85– 89	1
55–59	164	105–109	1
			1,000

6. Find the mean of the distribution of wages per day for employees of a certain factory.

Wages per Day Open limits	No. of Employees	Wages per Day Open Limits	No. of Employees
\$2.50–2.99	3	\$6.50–6.99	97
3 00–3 49	7	7.00–7.49	38
3.50–3.99	16	7.50–7.99	11
4.00–4.49	46	8 00–8.49	4
4 50–4.99	81	8 50–8.99	2
5.00–5.49	143	9.00–9.49	1
5 50–5.99	239	9.50–9.99	1
6.00–6.49	176		
			865

GRAPHICAL REPRESENTATION OF FREQUENCY DISTRIBUTIONS

The frequency distribution of petals on buttercups, as given on page 28, is plotted in Fig. 2.1. The horizontal axis represents the number of petals; the vertical, the number of flowers or frequencies.

FREQUENCY POLYGONS AND HISTOGRAMS

Figure 2.1 graphically represents the distribution of petals on these buttercups. Points b, c, d, etc., have been joined in Fig. 2.2 to make a frequency polygon *abcdefghijklma*. The object of this polygon is to enable one to see at once the nature of the frequencies. There were no buttercups with $8\frac{2}{3}$ petals, hence there is no

frequency at this point. Line *fg* excluding *f* and *g* represents no frequency. This polygon is used extensively to help the eye grasp the situation quickly, although it has no meaning for points between points plotted in Fig. 2.1.

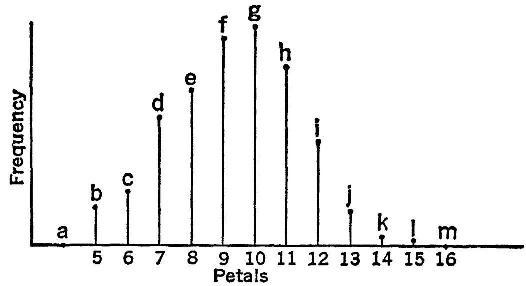


FIG. 2.1—Frequency distribution of petals on buttercups.

When data are grouped into classes, areas of rectangles are used to represent frequencies instead of vertical lines as in Fig. 2.1. The bases of these rectangles are equal to the width of the classes; the altitudes are equal to the respective frequencies. Each base

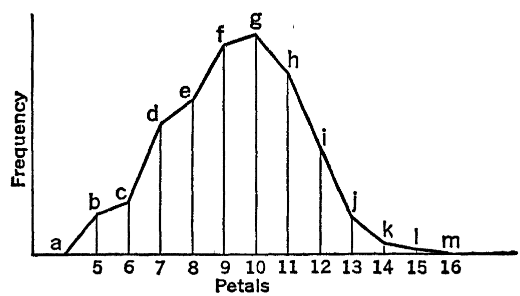


FIG. 2.2.—Frequency polygon of the distribution of petals on buttercups.

is equal to 1 class unit; hence the area is equal to the frequency. The representation of frequencies of the distribution of weights on page 31 is shown in Fig. 2.3.

Bases of the rectangles begin and end at the lower and upper closed class limits, respectively. Class marks are represented by points at the middle of the bases of the rectangles. The graph in which frequencies are represented by rectangles is called a histogram of the distribution. Figure *ABCDEFGHIJKA* is the his-

togram of the distribution mentioned. Rectangle CY represents the frequency or the number of students whose weights were between 109.75 pounds and 119.75 pounds. Rectangle PH represents the number whose weights were between 149.75 pounds and 159.75 pounds, etc. A histogram helps one to grasp immediately the nature of the frequencies. The vertical lines should always be drawn in, for they make the rectangles complete. These rect-

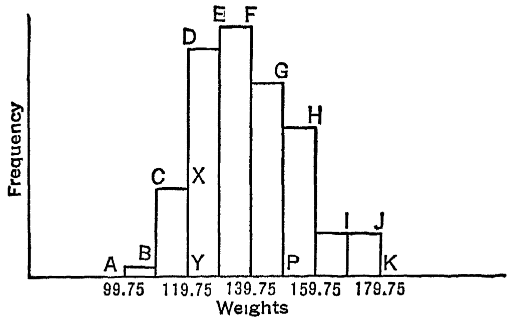


FIG. 2.3.—Histogram of the distribution of weights of male students.

angles are part of the histogram which represent frequencies; they should always be complete.

PROBLEMS

1. Represent graphically the distribution in problem 6 on page 29.

2. Represent graphically the distribution in problem 1 on page 24.

3. Represent graphically distributions in problems 4, 5, 6 on pages 35 and 36. Draw a vertical line at the mean of each. How much area is to the left of this line? How much is to the right of this line?

4. If a person picked at random a cluster from the plot from which the distribution in problem 5 on page 36 came, what is the probability of getting a cluster which had pedicels between 60 and 64 inclusive?

OGIVE

Figure 2.4 exhibits the total frequencies of Fig. 2.1 up to and including certain points. For example, the vertical line od in Fig. 2.4 gives the number of buttercups which had 5, 6, and 7 petals; line gp gives the number of buttercups which had 5, 6, 7, 8, 9, and 10 petals; etc. This graph of the cumulative frequencies is called an ogive or a cumulative frequency polygon. At the point

4 the ogive is zero, at 16 it is equal to 1,130. To the left of 5 it is zero, and to the right of 15 it is equal to 1,130. The ogive rises slowly at the left end of the range, rises rapidly for values in the

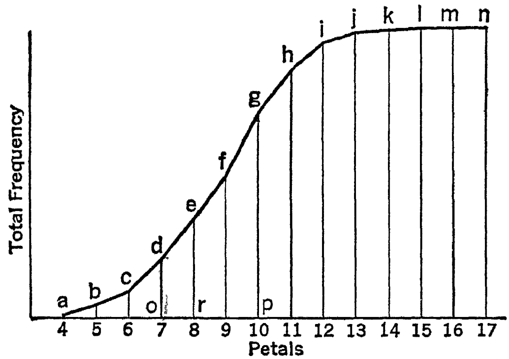


FIG. 2.4.—Ogive of the distribution of petals on buttercups.

middle of the range, and rises slowly at values near the right end of the range.

The ogive for the distribution in Fig. 2.3 is plotted in Fig. 2.5. Vertical line CX represents the total frequencies for students who

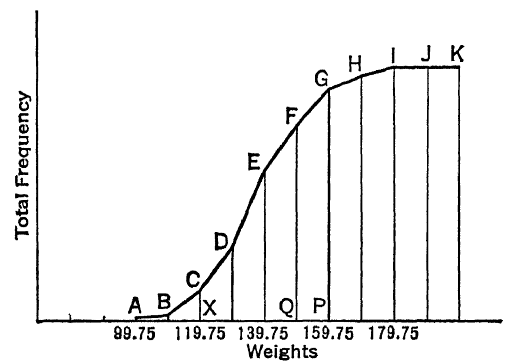


FIG. 2.5.—Ogive for the distribution of weights of male students.

weighed less than 120 pounds; line PG represents the number of students who weighed less than 160 pounds, etc. This ogive is equal to zero to the left of 99.75 pounds and equal to 100 to the right of 179.75 pounds. The difference between lines PG and QF gives the frequency for the class 149.75–159.75.

PROBLEMS

1. Construct the ogive for the distribution in problem 3 on page 25.

2. Construct the ogive for the distribution in the example concerning eggs laid by pheasants on page 27.

3. Construct the ogives for distributions in problems 4, 5, and 6 on pages 35 and 36.

4. From the ogive for the distribution in problem 5 on page 36 estimate the number of clusters which had 42 and 43 pedicels, the number which had 73, 74, 75, and 76 pedicels, and the number which had 51 pedicels.

5. From the ogive for the distribution in problem 6 on page 36 estimate the number of employees who received \$5.25–5.65, also the number who received \$6.15–6.40, also the number who received \$5.75.

THE MEDIAN

The median of a set of variates is the variate or quantity which has the same number of variates less than it as the number greater than it. The median of the variates, 4, 5, 9, 10, 11 is 9, as there are 2 variates less than 9 and 2 greater than 9. The median of the set 6, 8, 11, 12, 13, 16 is equal to any number between 11 and 12; it is customary to take the average of 11 and 12, or 11.5, as the median. In this case, the median is not one of the items. The median always divides the distribution into equal parts. A rule for finding the median for data which are not grouped is as follows:

(2.6) $m =$ the $\left(\dfrac{n+1}{2}\right)$th item, when the items are arranged in ascending or descending order.

In the first example the $\left(\dfrac{n+1}{2}\right)$th item is the $\left(\dfrac{5+1}{2}\right)$th item, or the third item. In the second example the $\left(\dfrac{n+1}{2}\right)$th item is the $\left(\dfrac{6+1}{2}\right)$th item, or the $3\frac{1}{2}$th item. There is no $3\frac{1}{2}$th item; it is considered to be the quantity half way between the third and fourth item, which is 11.5, as found before. Formula (2.6) always gives the median of a set of variates which are not grouped into classes.

PROBLEMS

1. Find the median of the distribution of the following scores made on a certain test: 80, 81, 76, 59, 55, 91, 63, 58, 82, 95, 79, 85, 83, 61, 58, 78, 82, 90, 86, 73, 59, 62, 71, 84, 87, 97, 67, 75, 86.

2. Find the median age of students in a class if the ages are: 19, 20, 18, 22, 25, 27, 21, 23, 24, 23, 26, 20.

MEDIAN FOR GROUPED DATA

Let it be required to find the median of the following distribution of weights of 123 men of age 20, where original measurements were made to the nearest 0.5 pound.

Weights, Classes	Frequency
99.75–109.75	3
109.75–119.75	11
119.75–129.75	27
129.75–139.75	32
139.75–149.75	23
149.75–159.75	17
159.75–169.75	6
169.75–179.75	4
	123

The median is the value which divides the area of the histogram into two equal parts; it is found by the formula

$$m = C_2 + \left(\frac{\frac{n}{2} - n_2}{f_2} \right) \cdot w, \tag{2.7}$$

where C_2 is the lower closed limit of the class in which the $\frac{n}{2}$th item lies, n_2 is the sum of all frequencies below C_2, and f_2 is the number of items in the class containing the $\frac{n}{2}$th item and w is the width of this class. The median usually lies in the class in which the $\frac{n}{2}$th item lies.

The $n/2$th item is the $61\frac{1}{2}$th item and is located in the fourth class. The lower limit of this class is $C_2 = 129.75$ pounds, $n_2 = 41$, $f_2 = 32$, and $w = 10$ pounds. The median is

$$m = 129.75 + \left(\frac{61.5 - 41}{32} \right) 10 = 129.75 + 6.41 = 136.16 \text{ lb.}$$

Line RS in Fig. 2.6 drawn through the point 136.16 pounds, perpendicular to the horizontal axis, has one half of the area to the left and one half to the right. For a grouped distribution the median is the value on the base line of the histogram, such that a perpendicular line to the base line through a point representing this value divides the area of the histogram into equal parts.

The median, in many cases, is a good representative of the items and is used by many investigators. Sizes of the items below and above the median do not affect its value, as the following

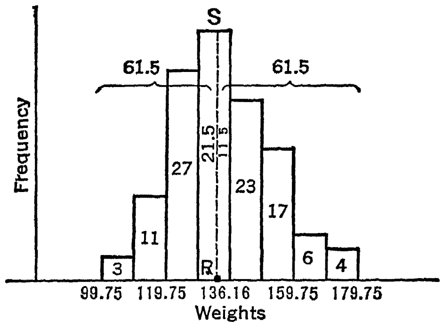


FIG. 2.6.—Distribution of weights of men 20 years of age, showing that the median divides the distribution into two equal parts.

illustration shows. The median of 3, 5, 6, 7, 9 is 6, and so is the mean. If 9 were changed to 14 the median would still be 6, while the mean would be changed to 7.

The median for grouped data cannot be obtained by formula (2.6) because a line perpendicular to the horizontal axis at the point obtained by this formula will not divide the area into two equal parts. This will be shown by finding the median as is done by some authors for the set of above weights. The formula which is given by some authors is

$$m = C_2 + \frac{\dfrac{n+1}{2} - n_2}{f_2} \cdot w. \quad \text{(Not correct for grouped data.)}$$

When this is used the median becomes

$$m = 129.75 + \frac{62 - 41}{32} 10 = 129.75 + 6.56 = 136.31 \text{ lb.}$$

If a line is drawn through this point perpendicular to the base line

in the histogram it does not divide the area into two equal parts. The area to the left is 62 and the area to the right is 61, which shows that the above formula is not the correct formula for finding the median of a grouped distribution. Formula (2.7) will give the median when the sum of frequencies is an odd or an even number.

The median is used as a representative of wages of employees of a factory when the mean would have no meaning. Let the wages per month of the employees be as follows:

Wages per Month
$ 75
85
105
115
125
130
140
200
600
$1,575

The mean is $175, and the median is $125. Here the median is a better representative of the wages than the mean. Values of extreme items greatly affect the mean but do not influence the median.

Let it be required to find the median of the distribution of children per family for a certain district.

No. of Children per Family	No. of Families
0	4
1	17
2	35
3	80
4	63
5	20
6	6
7	2
8	1
	228

Formula (2.7) must be used in this case to find the median, although the class width is unity. The median is

$$m = 2.5 + \frac{114 - 56}{80} \cdot 1 = 2.5 + 0.725 = 3.225 \text{ children.}$$

PROBLEMS

1. Find the median score made on a test in arithmetic if the scores were:

Scores	No. of Students	Scores	No. of Students
Below 39.5	7	69 5–74 5	189
39.5–44.5	17	74 5–79.5	261
44.5–49.5	26	79.5–84.5	119
49.5–54.5	38	84.5–89.5	80
54.5–59.5	53	89 5–94 5	18
59.5–64.5	79	94.5–99.5	3
64.5–69.5	104		994

Plot the data and erect a perpendicular at the median. Find the area to the left and also to the right of this line.

2. Find the median age for the city whose age distribution is as follows:

Age Group	No. of People	Age Group	No. of People
Under 1 year	2,190	25.29	9,833
1– 2	2,164	30–34	9,120
2– 3	2,326	35–44	17,198
3– 4	2,394	45–54	13,017
4– 5	2,368	55–64	8,396
5– 9	12,607	65–74	4,720
10–14	12,005	Over 75 years	1,911
15–19	11,552		122,671
20–24	10,870		

Can one find the mean of this set of data?

3. Find the median wage if wages per day are given below.

Wages	No. of Employees	Wages	No. of Employees
$2 00–2 24	4	$4 75–4.99	400
2.25–2 49	7	5.00–5.24	384
2.50–2 74	13	5.25–5.49	218
2.75–2.99	26	5.50–5.74	167
3.00–3 24	33	5.75–5.99	92
3.25–3.49	80	6 00–6.24	28
3.50–3 74	98	6 25–6.49	11
3.75–3 99	111	6.50–6.74	4
4.00–4.24	178		2,444
4.25–4 49	259		
4.50–4 74	331		

THE MODE

The mode of a set of variates is the variate which occurs the most frequently. In Fig. 2.1, 10 petals appeared more often than any other. In Fig. 2.3, the class mark of the fourth class might be taken as the approximate mode of weights; however, this class mark might not be the real mode. The mode of a grouped distribution is sometimes defined as the value of the variate at which the smooth curve which fits the data best has a maximum or greatest frequency. In some distributions there may not be a mode, for there may be two values at which the frequencies are greater than the others and are equal. The mode is sometimes spoken of as the typical value. Size 37 in men's suits is the mode, for more suits of this size are sold than of any other size.

There are various formulas for finding the approximate mode for grouped data. The following formula is used by many statisticians and is useful for distributions which have one class which has a frequency greater than any of the frequencies of the other classes; it is

$$\text{Mode} = L + \frac{f_2}{f_1 + f_2} \cdot w, \tag{2.8}$$

where L is the lower limit of the modal class or the class in which the mode falls, f_1 is the frequency of the class just below the modal class, f_2 is the frequency of the class just above the modal class, amd w is the width of the class.

Grouping has much to do with the value of the mode. Extreme values of the variates do not affect the value of the mode. The mode for the distribution on page 41 is

$$\text{Mode} = 129.75 + \frac{23}{27 + 23} \cdot 10 = 129.75 + 4.60 = 134.35 \text{ lb.}$$

PROBLEMS

1. Find the mode of the following distribution of bracts per cluster of *Daucus carota* (wild carrot). Plot the data, and compare the mean and median with the mode.

No. of Bracts per Cluster	No. of Clusters	No. of Bracts per Cluster	No. of Clusters
8	98	12	201
9	143	13	189
10	159	14	3
11	205	15	2

2. Find the mode of the heights of trees growing on a study plot if the following information was found by measurements after a certain lapse of time.

Height of Trees	No. of Trees
Under 5 ft.	49
5– 9	75
10–14	91
15–19	146
20–24	102
25–29	51
30–39	11
Over 40 ft.	3
	528

Compare the mode with the median.

3. Find the mode of the distribution of young per litter in mice (*Peromyscus truei truei* from Deadman Flat, near Flagstaff, Arizona):

No. in Litter	No. of Mother Mice
1	6
2	11
3	15
4	8
5	2
6	1
	43

THE GEOMETRIC MEAN

The geometric mean of a set of n positive variates is the nth root of their product. If the variates are $v_1, v_2, v_3, \ldots, v_n$, the geometric mean is by definition

$$G = \sqrt[n]{v_1 \cdot v_2 \cdot v_3 \cdot \ldots v_n}. \tag{2.9}$$

The logarithm of G is

$$\log G = \frac{1}{n} \Sigma \log v_i. \tag{2.10}$$

If the frequencies of the variates are respectively

$$f_1, f_2, f_3, \ldots, f_n,$$

the geometric mean is

$$G = \sqrt[\Sigma f]{v_1^{f_1} v_2^{f_2} v_3^{f_3} \ldots v_n^{f_n}}. \tag{2.11}$$

One of the objects of the geometric mean is to enable one to find averages of ratios and averages of rates of increase and decrease. Consider the following example. A certain city of 100,000 inhabitants increased from 1900 to 1910 by 6 per cent, from 1910 to 1920 by 9 per cent, and from 1920 to 1930 by 12 per cent; find the average rate of increase per decade. The 1930 census gave the population as 129,407. The required rate is the rate which will accumulate 100,000 into 129,407 in 3 conversion periods; that is, $100{,}000\ (1 + r)^3 = 129{,}407$, which is the ordinary compound interest law. The rate r must be the same for each decade.

Let us take the arithmetic average of the rates and determine whether it is the required rate. The average of the rates is 0.09. Using this rate for the average per decade, the population for 1910 is 109,000, that for 1920 is 118,810, and that for 1930 is 129,503, which is not equal to the population 129,407. This shows that the arithmetic mean of the rates is not the required rate.

The population in 1910 was 106 per cent of what it was in 1900; in 1920 it was 109 per cent of what it was in 1910, and in 1930 it was 112 per cent of what it was in 1920. The geometric mean of 1.06, 1.09, and 1.12 is

$$G = \sqrt[3]{(1.06)(1.09)(1.12)} = 1.08973,$$

and the average rate per 10-year period is 0.08973. According to this rate the population in 1910 should have been 108,973, in 1920 it should have been 118,751, and in 1930 it should have been 129,407, as it was.

The following data relate to values of an automobile.

Year	Value
1930	$720
1931	609
1932	500
1933	375
1934	227
1935	107

Find the average rate of decrease. The ratios of the value for 1 year divided by the value for the preceding year are:

$$\begin{aligned}
v_1 &= 609/720 = 0.84583; & \log(0.84583) &= 9.927{,}2831 - 10\\
v_2 &= 500/609 = 0.82102; & \log(0.82102) &= 9.914{,}3537 - 10\\
v_3 &= 375/500 = 0.75000; & \log(0.75000) &= 9.875{,}0613 - 10\\
v_4 &= 227/375 = 0.60533; & \log(0.60533) &= 9.781{,}9922 - 10\\
v_5 &= 107/227 = 0.47137; & \log(0.47137) &= \underline{9.673{,}3619 - 10}\\
& & & \ 49.172{,}0522 - 50
\end{aligned}$$

$$\log G = \frac{49.172{,}0522}{5} = 9.834{,}4105 - 10$$

$$G = 0.682984 = 1 - r,$$

where r is the rate of decrease. The average rate of decrease is 0.317016, for

$$720(1 - 0.317016)^5 = 107.$$

PROBLEMS

1. The following figures are quotations on a certain bond for 12 days: $54\frac{3}{8}$, $54\frac{5}{8}$, 55, $55\frac{1}{2}$, $55\frac{6}{8}$, $56\frac{1}{4}$, 56, $56\frac{1}{8}$, $55\frac{7}{8}$, $56\frac{3}{8}$, 57, $57\frac{1}{4}$. Find the average rate of increase per day. Plot these data.

2. The following table gives the number of children under 1 year of age out of 100,000 who were living during the first year for each month.

Months	No. Living Out of 100,000	Months	No. Living Out of 100,000
0–1	100,000	6–7	92,245
1–2	96,106	7–8	91,701
2–3	95,089	8–9	91,204
3–4	94,241	9–10	90,747
4–5	93,500	10–11	90,320
5–6	92,842	11–12	89,919

Find the average rate of decrease for the first 6 months and also for the last 6 months. Discuss the results.

3. The figures below are the population of whites and negroes of the United States of America from 1900 to 1930, inclusive.

Year	No. of Whites	No. of Negroes
1900	66,809,196	8,833,994
1910	81,364,447	9,827,763
1920	94,820,915	10,463,131
1930	108,864,207	11,891,143

Compare the average rate of increase per decade for whites and negroes.

4. Find the average rate of increase of the population of Canada. The population figures are:

Year	Population	Year	Population
1881	4,266,041	1911	7,169,960
1891	4,770,123	1921	8,767,206
1901	5,322,238	1931	10,376,786

5. Data below relate to exports from the United States before and during the depression:

	Before Depression		During Depression
Year	Exports	Year	Exports
1926	4,808,660,000	1930	3,843,181,000
1927	4,865,375,000	1931	2,424,289,000
1928	5,128,356,000	1932	1,611,016,000
1929	5,240,995,000	1933	1,674,975,000

Compare the average rates of increase or decrease for the two periods.

6. The geometric mean of 11 variates is 1.05; find the product of the variates.

7. The geometric mean of a set of variates is 1.079548, and the product of the variates is 2.1499. Find the number of variates in the set.

8. Plot: $A = 100(1 + r)^4$.

9. Plot: $A = 100(1 - r)^3$.

THE HARMONIC MEAN

The harmonic mean of a set of variates, $v_1, v_2, v_3, \ldots, v_n$, is equal to the number of the variates divided by the sum of the reciprocals of the variates, or

$$H = \frac{n}{\dfrac{1}{v_1} + \dfrac{1}{v_2} + \ldots + \dfrac{1}{v_n}} = \frac{n}{\Sigma \dfrac{1}{v}} \tag{2.12}$$

where $v_i \neq 0$.

If the variates have respectively the frequencies $f_1, f_2, \ldots, f_n$, then the harmonic mean of the set of variates is

$$(2.13) \qquad H = \frac{\Sigma f_i}{\dfrac{f_1}{v_1} + \dfrac{f_2}{v_2} + \ldots + \dfrac{f_n}{v_n}} = \frac{\Sigma f}{\Sigma \dfrac{f}{v}}$$

The following example illustrates the use of the harmonic mean.

Mr. Jones buys the articles listed below:

8 lb.	of sugar	for	$1 00,
10 "	of rice	"	1 00,
16 "	of beans	"	1 00,
6 "	of lard	"	1 00.

Find the average price per pound and the number of pounds he can buy at this average for $1.00.

1 lb. of	sugar	costs	$\$\frac{1}{8}$,
" " "	rice	"	$\$\frac{1}{10}$,
" " "	beans	"	$\$\frac{1}{16}$,
" " "	lard	"	$\$\frac{1}{6}$,

hence the average price per pound is

$$\frac{\frac{1}{8} + \frac{1}{10} + \frac{1}{16} + \frac{1}{6}}{4} = \$0.1136.$$

The number of pounds that can be bought at this average price is the harmonic mean of the number of pounds purchased, and this is

$$\frac{1}{\dfrac{\frac{1}{8} + \frac{1}{10} + \frac{1}{16} + \frac{1}{6}}{4}} = \frac{1}{0.1136} = 8.80 \text{ lb.}$$

The harmonic mean in the above example gave the number of pounds which could be purchased at the average price for $1.00.

Consider that a man travels in his car for 30 minutes at 40 miles per hour, during the next 20 minutes at 30 miles per hour, and during the following 10 minutes at 20 miles per hour. What was his rate per hour?

He goes	$(30/60)40 = 20$	mi. during the first	30 min.,
" "	$(20/60)30 = 10$	" " " next	20 " ,
" "	$(10/60)20 = 3.33$	" " " "	10 " ,

hence in the hour he traveled 33.33 miles, and this is his rate per hour. This is a harmonic mean, for his rate is distance divided by time and is

$$\frac{20 + 10 + 3.33}{\frac{20}{40} + \frac{10}{30} + \frac{3\ 33}{20}} = 33\ 33 \text{ mi. per hr.}$$

The following was observed with regard to mowing a large university campus:

5	men	mowed	the	campus	in	3 days,
3	"	"	"	"	"	6 " ,
7	"	"	"	"	"	2 " .

On the average how long would it take 1 man to mow the campus?

Each	of	the	first	5	men	mowed	$\frac{1}{15}$th	of	the	campus	per day,
"	"	"	"	3	"	"	$\frac{1}{18}$th	"	"	"	" "
"	"	"	"	7	"	"	$\frac{1}{14}$th	"	"	"	" "

The average amount of the campus that was mowed per day by 1 man was

$$\frac{\frac{5}{15} + \frac{3}{18} + \frac{7}{14}}{15} = 0.067\text{th of the campus per day.}$$

The length of time it would take 1 man on the average to mow the lawn would be the harmonic mean

$$\frac{15}{\frac{5}{15} + \frac{3}{18} + \frac{7}{14}} = 15 \text{ days.}$$

PROBLEMS

1. The following is the record of an airplane for 3 hours:

For	the	first	80	min.	it	traveled	at	the	rate	of	150	mi.	per	hr.,
"	"	next	40	"	"	"	"	"	"	"	130	"	"	",
"	"	"	30	"	"	"	"	"	"	"	100	"	"	",
"	"	"	20	"	"	"	"	"	"	"	60	"	"	",
"	"	"	10	"	"	"	"	"	"	"	50	"	"	".

Find the average speed per hour.

2. Mr. Henry purchased the following groceries:

7 lb. of lard for $1.00,	14 lb. of bread for $1.00,
4 " " butter for $1.00,	180 " " salt " $2.00,
5 " " coffee " $1.00,	25 " " sugar " $1.25.

Find the average cost per pound and the average number of pounds at this average that can be purchased for $1.00.

3. If 9 men can do a piece of work in 6 days,

4	" " " the same piece of work in	10 days,
6	" " " " " " " " "	7 " ,
12	" " " " " " " " "	4 " .

Find the time it takes on the average for 5 men to do the piece of work and the length of time it takes the average man to do the work.

4. A store has:

50 Latin	books at ..	$2.10	per book,
25 Greek	" "	1.85	" " ,
35 German	" "	1.25	" " ,
17 Spanish	" "	1.65	" " ,
11 English	" "	.95	" " ,
43 French	" "	1.30	" " ,
27 Russian	" "	2.80	" " .

Find the average price per book.

5. A machine requiring 3 men to operate makes 50 articles in 40 min.,

Another	" 5 " " "	" 62 "	" 30 " ,
"	" 2 " " "	" 30 "	" 20 " .

Find the average time it takes the machines to make the article. How many men are required to make 1 article per minute?

6. For a set of positive variates prove that

$$M_v \geqq G_v \geqq H_v.$$

7. If the harmonic mean of n variates is 240/57 and the sum of the reciprocals of the variates is 57/60, find n.

8. If the harmonic mean of 10 items is 25, find the sum of the reciprocals of the variates.

CHAPTER 3

MEASURES OF DISPERSION

MEAN DEVIATION

Each of two marksmen A and B stands at the same place and shoots 10 shots at a horizontal line CE drawn on the target wall. Figures 3.1 and 3.2 exhibit their records.

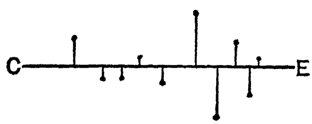


FIG. 3.1.—A's record.

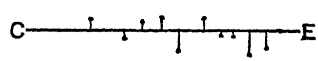


FIG. 3.2.—B's record.

The average of the distances of the shots from line CE gives a good idea how close the shots came to the line or how far the shots fell from the line. The average of the deviations all taken positive for A is 8 5 mm.; this average for B is 3.9 mm. Clearly B is the better marksman. If the average of the deviations of the shots from CE is small the shots fell close to the line; if this average is large the shots fell farther from the line. Marksmen can be compared by comparing the averages of the deviations of the shots from the line at which they shoot.

In a similar way the average of the deviations of variates from the mean of the variates all taken positive shows how near the items are to the mean. If the average of the absolute values (absolute value means the value taken positive) of the deviations of the variates from the mean is large then there is a wide scattering of the variates about the mean; if this value is small the variates differ very little from the mean and the mean is a better representative of the central tendency of the variates than when the average of the deviations is large.

The average of the absolute values of the deviations of the

variates $v_1, v_2, v_3, \ldots, v_n$ from the mean of the variates is called the mean deviation of the set of variates and is

$$\text{(3.1)} \qquad \text{M.D.} = \frac{\Sigma \mid v_i - M_v \mid}{n} = \frac{\Sigma \mid \bar{v} \mid}{n},$$

where $\bar{v}_i$ represents the deviation of the ith item from the mean. If the items form a frequency distribution, the mean deviation of the items is

$$\text{(3.2)} \qquad \text{M.D.} = \frac{\Sigma \mid v_i - M \mid \cdot f(v_i)}{\Sigma f(v_i)} = \frac{\Sigma \mid \bar{v}_i \mid \cdot f(\bar{v}_i)}{\Sigma f(\bar{v}_i)}\,; [f(v_i) = f(\bar{v}_i)].$$

The following measurements are lengths of feet of men 20 years of age.

Length of Feet in Inches v	$\bar{v}$	$\bar{v}^2$	Length of Feet in Inches v	$\bar{v}$	$\bar{v}^2$
11.8	+0.04	0.0016	11.6	+0.05	0 0025
11.3	−0.46	0.2116	11.4	−0.15	0.0225
11.9	+0 14	0.0196	11.5	−0.05	0 0025
12.4	+0.64	0.4096	11 1	−0.45	0.2025
11.1	−0 66	0.4356	11 6	+0.05	0.0025
12 6	+0.84	0.7056	11 3	−0.25	0 0625
10 9	−0 86	0.7396	12.0	+0.45	0.2025
11.5	−0 26	0.0676	11 6	+0 05	0 0025
12.0	+0.24	0.0576	11 5	−0.05	0.0025
12.1	+0 34	0 1156	11.9	+0.35	0 1225
117.6	0.00	2.7640	115.5	0.00	0 6250

$$\text{M.D.} = \frac{\Sigma \mid \bar{v} \mid}{10} = \frac{4.48}{10} = 0.448 \text{ in.}; \ \text{M.D.} = \frac{\Sigma \mid \bar{v} \mid}{10} = \frac{1.90}{10} = 0.19 \text{ in.}$$

Items in the second set of measurements lie closer to their mean than items in the first set to their mean. The mean deviation is a measure of dispersion, for it shows how the items are dispersed or scattered about the mean. A small mean deviation signifies little scattering of the variates about the mean; a large mean deviation denotes extensive scattering of the variates about the mean.

Consider the distribution of terminal inflorescences of water hemlock (*Cicuta maculata*).

No. of Terminal Inflorescences per Plant v	No. of Plants $f(v)$	$vf(v)$	$\bar{v}$	$\lvert\bar{v}\rvert f(\bar{v})$ *	$\bar{v}f(\bar{v})$
3	8	24	−2	16	−16
4	19	76	−1	19	−19
5	44	220	0	0	0
6	24	144	+1	24	24
7	4	28	+2	8	8
8	1	8	+3	3	3
	100	500		70	0

* $f(\bar{v}) = f(v)$.

The mean and mean deviation are respectively:

$$M_v = 5 \text{ terminal inflorescences per plant,}$$

$$\text{M.D.}_v = \frac{\Sigma \mid \bar{v} \mid f(v)}{\Sigma f(v)} = \frac{70}{100} = 0.70 \text{ terminal inflorescence.}$$

The mean deviation can be found without finding the deviations of the variates from the mean. If a deviation of a variate v_i from the mean, that is, $v_i - M$, is negative, then $-(v_i - M)$ is positive. According to this the sum of all possible deviations minus the sum of all negative deviations divided by n is M.D., or

$$\text{(3.3)} \qquad \text{M.D.} = \frac{\overset{+}{\Sigma}(v_i - M) - \overset{-}{\Sigma}(v_i - M)}{n},$$

where $\overset{+}{\Sigma}$ designates the sum of all positive deviations from the mean and $\overset{-}{\Sigma}$ designates the sum of all negative deviations from the mean. Equation (3.3) may be written as:

$$\text{(3.4)} \quad \text{M.D.} = \frac{\overset{+}{\Sigma}v_i - n_1M - \overset{-}{\Sigma}v_j + n_2M}{n} = \frac{\overset{+}{\Sigma}v_i - \overset{-}{\Sigma}v_j - (n_1 - n_2)\cdot M}{n}$$

where n_1 is the number of variates greater than the mean and n_2 is the number of variates less than the mean. The number n is not always equal to $n_1 + n_2$ because some of the deviations from the mean may be zero.

Formula (3.4) enables one to find the mean deviation without finding deviations from the mean, for it is in terms of the original variates, the number n_1 of variates greater than the mean, n_2 the number less than the mean, M the mean, and n the number of

variates. According to the above formula the mean deviation of the first set of feet lengths of page 54 is

$$\text{M.D.} = \frac{72.8 - 44.8 - (6 - 4)11.76}{10} = \frac{28.0 - 23.52}{10} = 0.448 \text{ in.}$$

When the variates are in a frequency distribution the formula for the mean deviation is

$$(3.5) \qquad \text{M.D.} = \frac{\overset{+}{\Sigma} v_i f(v_i) - \overset{-}{\Sigma} v_j f(v_j) - [\overset{+}{\Sigma} f(v_i) - \overset{-}{\Sigma} f(v_j)] \cdot M_v}{\Sigma f(v)},$$

where the summation $\overset{+}{\Sigma} v_i f(v_i)$ denotes the sum of all variates greater than the mean and $\overset{-}{\Sigma} v_j f(v_j)$ denotes the sum of those less than the mean. If a computing machine is available, it is easy to find the mean deviation, for one turns the crank forward for variates greater than the mean and backward for those less than the mean; this gives the difference of the first two terms in (3.4) and (3.5). By noticing the counter in the upper dial of numbers on the machine, the value of the quantity in parentheses or brackets in these formulas is at once obtained. From these values one can readily find M.D. This is by far the easiest method, for it is not necessary to use paper and pencil for any of the computations.

In all these examples the algebraic sum of the deviations of the items from the mean is zero. This is always true, as the next theorem will show.

THEOREM 3.1. The algebraic sum of the deviations of the variates from the mean of the set of variates is zero.

PROOF: By definition the deviations from the mean are:

$$\begin{aligned}
\bar{v}_1 &= v_1 - M_v \\
\bar{v}_2 &= v_2 - M_v \\
\bar{v}_3 &= v_3 - M_v \\
&\dots\dots\dots\dots \\
\bar{v}_n &= v_n - M_v
\end{aligned}$$

$$\text{Adding: } \Sigma\bar{v} = \Sigma v - nM_v = \Sigma v - n(\Sigma v/n) = \Sigma v - \Sigma v = 0.$$

If one finds M.D. by calculating the deviations from the mean, a check on the computations is that the sum of these deviations

should be zero, or approximately zero, if deviations are correct to certain decimal places.

INDIRECT METHOD OF CALCULATING M.D.

Suppose h (= 11.5 inches) is subtracted from each item of the first set on page 54. Results from these subtractions are listed in the next table. Column d contains a set of items much smaller than the original items. It has been shown that the mean of the v's can be found by using the d's. It will now be shown that the M.D. of the v's can be found by using the d's. Set up the $d - M_d$

Length of Feet v	$v - h$ d	$v - M_v$	$d - M_d$
11 8	+0 3	+0.04	+0.04
11 3	−0.2	−0.46	−0 46
11.9	+0.4	+0 14	+0.14
12 4	+0.9	+0.64	+0.64
11.1	−0.4	−0 66	−0.66
12.6	+1.1	+0 84	+0 84
10.9	−0 6	−0.86	−0.86
11.5	0.0	−0.26	−0 26
12.0	+0 5	+0.24	+0.24
12.1	+0.6	+0.34	+0.34

$M_v = 11.76$ in.; $M_d = 0.26$.

column, or the set of deviations of the d's from the mean of the d's. By comparing columns $v - M_v$ and $d - M_d$ it is seen that the deviations of the v's from the mean of the v's are identically equal to the corresponding deviations of the d's from the mean of the d's. This example shows that for this case M.D.$_{\cdot v}$ = M.D.$_{\cdot d}$.

Theorem 3.2. From each variate of a set of variates subtract a constant h, thus forming a new set of variates. The deviation of the ith variate in the original set from the mean of the original variates is equal to the deviation of the ith variate in the new set from the mean of the new set of variates, or

$$v_i - M_v = d_i - M_d, \quad \text{or} \quad \bar{v}_i = \bar{d}_i.$$

Proof: By definition

$$d_i = v_i - h.$$

By formula (2.2), $h = M_v - M_d$. Substituting this value for h in the above equation gives

$$d_i = v_i - M_v + M_d, \quad \text{or} \quad d_i - M_d = v_i - M_v, \quad \text{or} \quad \bar{d}_i = \bar{v}_i,$$

which proves the theorem.

This theorem enables one to set up a new set of variates which are smaller than the original set and enables one to find the mean deviation of the original set by finding the mean deviation of the new set of smaller variates.

The mean deviation of a grouped distribution is approximately equal to the mean deviation of the distribution made up of class marks.

PROBLEMS

1. Find the mean deviation of the following set of items by finding each deviation from the mean.

No. of Auto Tires Sold per Day	No. of Auto Tires Sold per Day
21	26
23	17
19	24
20	22
18	16

How many items are within 1 M.D. of the mean?

2. Find the mean deviation by use of a provisional mean and the number of variates within 1 M.D. of the mean of the distribution of rows of kernels on ears of corn:

No. of Rows of Kernels on Ears of Corn	No. of Ears of Corn, Frequency
10	2
12	32
14	218
16	482
18	470
20	232
22	82
24	20
	1,538

3. Find the mean deviation, the percentage of items within 1 M.D. of the mean, and the percentage within 2 M.D. of the mean, of the following distribution:

No. of Alpha Particles Radiates From a Disk in $\frac{1}{8}$ Minute *	Frequency
0	57
1	203
2	382
3	525
4	532
5	408
6	273
7	139
8	45
9	27
10	10
11	4
12	0
13	1
14	1
	2,607

THE STANDARD DEVIATION

Square roots of the averages of the squares of deviations of shots from line CE in Figs. 3.1 and 3.2 also measure the amount of scattering or dispersion of the shots about CE. The square root of the average of the squares of the deviations for marksman A is about 9.1 mm.; that for marksman B is about 4.4 mm. The larger value indicates the greater scattering of shots about CE and hence reveals the poorer marksman.

In the same way the square root of the average of the squares of deviations of variates from the mean of the variates is a measure of dispersion or a measure of the scattering of the items about the mean. The smaller this quantity is, the less amount of scattering; if there is little scattering then the mean is a good representative of the central tendency of the items.

The standard deviation of a set of variates is the square root of

* Rutherford and Geiger, "The Probability Variation in the Distribution of Alpha Particles," *Phil. Mag.*, Series 6, Vol. 20 (1910), pp. 698–701.

the average of the squares of the deviations of the variates from the mean of the variates, or

$$\text{(3.6)} \qquad \text{S.D.} = \sigma_v = \sqrt{\frac{\Sigma(v - M_v)^2}{n}} = \sqrt{\frac{\Sigma\bar{v}^2}{n}}$$

where σ_v is the symbol for the standard deviation, and $\bar{v}_i = v_i - M_v$.

When variates form a frequency distribution the standard deviation is

$$\text{(3.7)} \qquad \text{S.D.} = \sigma_v = \sqrt{\frac{\Sigma(v_i - M)^2 f(v_i)}{\Sigma f(v_i)}} = \sqrt{\frac{\Sigma\bar{v}^2 f(\bar{v})}{\Sigma f(\bar{v})}}.$$

Consider again the length of feet given on page 54. The sum of the squares of the deviations of the lengths from the means are given on page 54. From these values the standard deviations are respectively

$$\sigma_1 = \sqrt{0.2764} = 0.53 \text{ in}, \qquad \sigma_2 = 0.0625 = 0.25 \text{ in}.$$

The standard deviation of the first set of lengths is 0.53 inch and is larger than the standard deviation of the second set, because there is more scattering of the items in the first set about the mean than there is in the second. The standard deviation is used more often in statistics than any other measure of dispersion.

The standard deviation is roughly about 125 per cent of the mean deviation.

Formula (3.6) can be written in terms of the original variates instead of the deviations from the mean as :

$$\text{(3.8)} \qquad \sigma_v = \sqrt{\frac{\Sigma(v - M)^2}{n}} = \sqrt{\frac{\Sigma(v^2 - 2v \cdot M + M^2)}{n}}$$

$$= \sqrt{\frac{\Sigma v^2}{n} - \frac{2M \cdot \Sigma v}{n} + \frac{nM^2}{n}} = \sqrt{\frac{\Sigma v^2}{n} - 2 \cdot M^2 + M^2}$$

$$= \sqrt{\frac{\Sigma v^2}{n} - M^2} = \sqrt{\frac{\Sigma v^2}{n} - \left(\frac{\Sigma v}{n}\right)^2}.$$

In a similar way formula (3.7) may be written in terms of frequencies

$$\text{(3.9)} \qquad \sigma_v = \sqrt{\frac{\Sigma(v - M)^2 f(v)}{\Sigma f(v)}} = \sqrt{\frac{\Sigma v^2 f(v)}{\Sigma f(v)} - \left(\frac{\Sigma v f(v)}{\Sigma f(v)}\right)^2}.$$

Formulas (3.8) and (3.9) express the standard deviation in terms of the average of the squares of the original items and the average of the items; hence it is not necessary to determine deviations from the mean.

Consider the distribution of petals of buttercups:

No. of Petals v	Frequency $f(v)$	$v \cdot f(v)$	$v^2 f(v)$	$(v+1)^2 f(v)$
5	40	200	1,000	1,440
6	52	312	1,872	2,058
7	126	882	6,174	8,064
8	165	1,320	10,560	13,385
9	204	1,836	16,524	20,400
10	215	2,150	21,500	26,015
11	177	1,947	21,417	25,488
12	104	1,248	14,976	17,576
13	35	455	5,915	6,860
14	8	112	1,568	1,800
15	4	60	900	1,024
	1,130	10,522	102,406	124,580

$M_v = 9.3115^+$ petals.

$$\sigma_v = \sqrt{\frac{\Sigma\, v^2 f(v)}{\Sigma\, f(v)} - \left(\frac{\Sigma\, v \cdot f(v)}{\Sigma\, f(v)}\right)^2} = \sqrt{90.6248 - 86.7040}$$

$$= \sqrt{3.9208} = 1.9801 \text{ petals.}$$

It is essential to check all computations to avoid errors which might arise. The last column in the above table enables one to check computations. The summation of the last column expanded is

$$\Sigma(v+1)^2 f(v) = \Sigma(v^2 + 2v + 1)f(v) = \Sigma v^2 f(v) + 2\Sigma v f(v) + n.$$

This means that the sum of the quantities in the last column should be equal to the sum of the fourth column plus twice the

sum of the third column plus the sum of the second. For the above table these values are

$$124{,}580 = 102{,}406 + 2(10{,}522) + 1{,}130 = 124{,}580,$$

which check the computations.

It is of great importance to know the number of items within 1, 2, 3, and 4 standard deviations of the mean. For the distribution of petals these values are as follows:

Number	of	items	within	$1\ \sigma$	of	the	mean	is	761 or 67.3%,
"	"	"	"	$2\ \sigma$	"	"	"	"	1,078 or 95.4%,
"	"	"	"	$3\ \sigma$	"	"	"	"	1,130 or 100%.

Nearly two-thirds of the variates are within 1 σ of the mean. All the variates are within 3 σ's of the mean. In many distributions found from observations, about two-thirds of the variates lie within 1 σ of the mean; about 95 per cent lie within 2 σ's of the mean, and practically all lie within 3 σ's of the mean. In a large majority of distributions very few variates lie beyond 4 σ's of the mean. These facts indicate the importance of the mean and standard deviation of a distribution.

Since the standard deviation is the square root of the average of the squares of deviations of variates from the mean it is the same as the standard deviation of a set of variates formed by subtracting a provisional mean from each variate in the distribution. See theorem 3.2.

Formula (3.9) can now be written

$$(3.10) \qquad \sigma_v = \sqrt{\frac{\Sigma(v - M_v)^2 f(v)}{\Sigma f(v)}} = \sqrt{\frac{\Sigma(d - M_d)^2 f(d)}{\Sigma f(d)}}$$

$$= \sqrt{\frac{\Sigma d^2 f(d)}{\Sigma f(d)} - \left(\frac{\Sigma d f(d)}{\Sigma f(d)}\right)^2} = \sigma_d.$$

This will be illustrated by the distribution of petals of buttercups on page 61. Let the provisional mean $h = 10$ petals. Subtract 10 from each variate, and place these values in column 3 of the following table.

No. of Petals	$f(v)$	d	$df(d)$	$d^2f(d)$	$(d+1)^2f(d)$
5	40	-5	$-$ 200	1,000	
6	52	-4	$-$ 208	832	
7	126	-3	$-$ 378	1,134	
8	165	-2	$-$ 330	660	
9	204	-1	$-$ 204	204	
10	215	0	$-1{,}320$	000	
11	177	$+1$	177	177	
12	104	2	208	416	
13	35	3	105	315	
14	8	4	32	128	
15	4	5	20	100	
	1,130		+ 542	4,966	4,540
			$-$ 778		

$$M_d = -\ 0.6885^{-} \text{ petal.}$$

From (3.10)

$$\sigma_v = \sqrt{4.3947 - 0.4740} = 1.9801 \text{ petals.}$$

As a check on computations the sum of the sixth column should be equal to the sum of the fifth column plus twice the sum of the fourth plus the sum of the second column. These values are

$$4{,}540 = 4{,}966 + 2(-778) + 1{,}130 = 4{,}540,$$

which check computations in the table.

Formula (3.8) may now be written as

$$\sigma_v = \sqrt{\frac{\Sigma(v - M_v)^2}{n}} = \sqrt{\frac{\Sigma(d - M_d)^2}{n}} \tag{3.11}$$

$$= \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2}.$$

It is not necessary to write quantities in columns 4, 5, and 6 if one has access to a computing machine, for one can readily sum these columns on the machine by the cumulating process.

Formulas (3.10) and (3.11) are employed by most workers in statistics because they allow the use of a provisional mean, which enables one to reduce the size of the variates and greatly diminish the amount of computing.

Formulas (3.10) and (3.11) may be written as follows:

$$(3.12)\quad \text{S.D.} = \sigma = \sqrt{\frac{\Sigma d^2}{n} - \left(\frac{\Sigma d}{n}\right)^2} = \frac{1}{n}\sqrt{n\Sigma d^2 - (\Sigma d)^2};$$

$$(3.13)\quad \text{S.D.} = \sigma = \sqrt{\frac{\Sigma d^2 f(d)}{\Sigma f(d)} - \left[\frac{\Sigma d f(d)}{\Sigma f(d)}\right]^2}$$

$$= \frac{1}{\Sigma f(d)}\sqrt{\Sigma f(d)\cdot \Sigma d^2 f(d) - [\Sigma d\cdot f(d)]^2}.$$

H. C. Carver gave the above formulas for the first time and likes to use them in preference to others. (See "Elementary Mathematical Statistics," by H. C. Carver.) A disadvantage of formulas (3.12) and (3.13) is that they involve such large numbers. According to formula (3.13) the standard deviation of the distribution on page 63 is

$$= \frac{1}{1{,}130}\sqrt{1{,}130(4{,}966) - (778)^2} = \frac{1}{1{,}130}\sqrt{5{,}006{,}296}$$

$$= \frac{2{,}237.4754}{1{,}130} = 1.9801 \text{ petals, as found before.}$$

MOMENTS OF A DISTRIBUTION

The nth moment of a frequency distribution about a fixed point* (usually about zero) is equal to the average of the nth powers of the variates measured from this fixed point, or

$$(3.14)\qquad \mu'_{n.v} = \frac{\Sigma v^n}{n} = \frac{\Sigma v^n f(v)}{\Sigma f(v)}.$$

The nth moment of a distribution of variates about the mean of the variates is the average of the nth powers of the deviations of the variates from the mean, or

$$(3.15)\qquad \mu_{n:v} = \frac{\Sigma(v - M_v)^n}{n} = \frac{\Sigma(v - M_v)^n f(v)}{\Sigma f(v)} = \frac{\Sigma \bar{v}^n f(v)}{\Sigma f(v)}.$$

According to these definitions the mean of a set of variates is the first moment, the average of the squares is the second moment, etc. The first two moments in both cases are

* The moment about a fixed point a is $\underset{a}{\mu'_{n:v}} = \dfrac{\Sigma(v - a)^n}{n}$.

First moment

$$\mu'_{1\cdot v} = M_v = M = \frac{\Sigma v f(v)}{\Sigma f(v)}\text{ ; second moment } \mu'_{2\,.\,v} = \frac{\Sigma v^2 f(v)}{\Sigma f(v)}.$$

First moment about the mean is

$$\mu_{1:v} = \frac{\Sigma(v - M)f(v)}{\Sigma f(v)} = \frac{\Sigma \bar{v}\cdot f(\bar{v})}{\Sigma f(\bar{v})} = 0.$$

Second moment about the mean $\mu_{2\ v} = \dfrac{\Sigma(v - M)^2 f(v)}{\Sigma f(v)} = \dfrac{\Sigma \bar{v}^2 \cdot f(\bar{v})}{\Sigma f(\bar{v})}$.

The second moment about the mean may be written as follows:

$$\mu_{2:v} = \frac{\Sigma(v - M)^2 f(v)}{\Sigma f(v)} = \frac{\Sigma(v^2 - 2v\cdot M + M^2)f(v)}{\Sigma f(v)}$$

$$= \frac{\Sigma v^2 f(v)}{\Sigma f(v)} - \frac{2\Sigma f(v) v\cdot M}{\Sigma f(v)} + \frac{\Sigma f(v)\cdot M^2}{\Sigma f(v)}$$

$$= \mu'_{2:v} - 2M^2 + M^2 = \mu'_{2\,.\,v} - M^2,$$

or

$$\mu_{2:v} = \mu'_{2:v} - M^2. \tag{3.16}$$

PROBLEMS

1. A stem of a certain water plant was broken off and inverted in water. Tiny bubbles passed out of the end that was cut. Layers of waxed paper were held between the sun and the plant. The following table gives the distribution of the number of bubbles per minute.

Layers of Wax Paper	No. of Bubbles per Minute	Layers of Wax Paper	No. of Bubbles per Minute	Layers of Wax Paper	No. of Bubbles per Minute
0	36	9	56	18	32
1	38	10	58	19	26
2	39	11	57	20	21
3	41	12	55	21	16
4	43	13	54	22	9
5	46	14	51	23	5
6	47	15	47	24	2
7	49	16	44	25	1
8	53	17	38		
					964

Find the number of layers or the light intensity which produced the most bubbles and the standard deviation of the distribution.

2. Find the number of items within 1, 2, and 3 standard deviations of the mean of the distribution in problem 2 on page 28 and express in percentages.

3. Find the percentage of the items within 1, 2, and 3 σ's of the mean of the distribution in problem 1 on page 65.

4. The standard deviation of a distribution is 1.63 rays, and the mean is 9.93 rays; find the second moment of the distribution.

5. The second moment of a set of variates is 24.6 square gallons; the standard deviation is 0.47 gallon; find the mean of the distribution.

6. Prove that $\Sigma\,(v - M)^2 = \Sigma\,v^2 - (\Sigma\,v)^2/n$.

7. The mean and standard deviation of 100 items are respectively 80 people and 5 people. If the 100th item, 86 people, is omitted from the data, find the sum of the squares of the 99 items.

8. Using the data in problem 7 find the standard deviation of the 99 items.

9. If the mean deviation of a set of items is 7.8 balls and the sum of the deviations of the items from the mean taken positive is 3,900 balls, find the number of items.

THE STANDARD DEVIATION OF GROUPED DATA

When variates are grouped into classes, individual variates going into a class are represented by the class mark. If the variates are uniformly distributed, this class mark is a good representative of the variates in the class, and the sum of the variates in the class is equal to the class mark times the frequency (number in the class) of the class. The sum of the squares of the variates in a class is not equal to the class mark squared multiplied by the number of variates in the class. Suppose that a class from 6.5 to 9.5 had the following variates before the grouping was made.

	v	$f(v)$	$vf(v)$	$v^2f(v)$
	7	1	7	49
Class{6.5 − 9.5}	8	2	16	128
	9	1	9	81
Class mark 8		4	32	258

The mean of the items in this class is equal to the class mark 8, and the sum of the items in this class, 32, is equal to the class mark,

8, times the frequency 4 of the class; but the sum of the squares, 258, is not equal to the class mark squared, times the frequency of the class. In the above class the sum of the squares of the items is 258, while the square of the class mark times the frequency, or number in the class, is 256, a difference of 2 square units. To compensate for similar differences throughout the distribution for all classes, a correction is made, which enables one to find the standard deviation approximately for the grouped data. In advanced courses in statistics it has been shown that the approximate standard deviation for a grouped distribution of discrete variates is

$$\sigma_{\text{adj.}} = \left(\sqrt{\mu'_{2:d} - M_d^2 - \frac{1 - \frac{1}{k^2}}{12}}\right)\cdot w, \tag{3.17}$$

where k is the number of values of the variates that can go into a class, and where discrete variates are variates which differ by whole units.* Things which can be counted are discrete variables, for example, the number of children in a family, the number of petals on flowers, the number of stamps sold by a post office, the number of cents paid employees, etc. In the above class, $k = 3$, for 7, 8, and 9 are the only discrete variates that can fall in this class, 6.5–9.5. In problem 2 on page 35, $k = 2$; in problem 5 on page 36, $k = 5$.

The following distribution of pedicels per inflorescence of a certain plant will be used to show that it is necessary to make adjustments in order to obtain approximately the correct standard deviation of the distribution of grouped data. The mean and standard deviation of the distribution were obtained before the variates were grouped into classes. For each grouping the mean, unadjusted σ and the adjusted σ were obtained. The adjusted standard deviations were determined from formula (3.17). The distribution is given below for the data before grouping.

* $\sigma_{\text{adj.}}$ = adjusted standard deviation.

PEDICELS OF *Daucus carota* (WILD CARROT)

Pedicels per Cluster	No. of Clusters	Pedicels per Cluster	No. of Clusters	Pedicels per Cluster	No. of Clusters
12	2	41	386	70	62
13	3	42	380	71	87
14	4	43	360	72	59
15	6	44	380	73	47
16	8	45	349	74	52
17	8	46	367	75	26
18	9	47	332	76	33
19	28	48	355	77	37
20	15	49	390	78	24
21	19	50	340	79	37
22	43	51	423	80	18
23	54	52	360	81	20
24	57	53	367	82	17
25	74	54	362	83	16
26	97	55	325	84	7
27	121	56	309	85	8
28	156	57	275	86	8
29	178	58	236	87	6
30	203	59	258	88	6
31	221	60	216	89	4
32	243	61	211	90	3
33	285	62	188	91	2
34	247	63	169	92	2
35	296	64	141	93	6
36	288	65	113	94	3
37	305	66	128	95	1
38	350	67	129	96	2
39	288	68	115	97	1
40	408	69	126		
					12,800

Correct values for M and σ are those that were found before grouping; these values are listed in the first line of the table below. Results for various ways of grouping are given also in the table. The first group in every case had 11.5 as lower limit.

Groups of	Means	Unadjusted σ	Adjusted σ
1	47 3060	12.6711	12 6711
2	47.2984	12.6755	12.6656
3	47.3104	12.6972	12 6710
4	47.2972	12.7164	12.6671
5	47.2977	12.7473	12.6686
6	47.3003	12 8006	12.6860
7	47 3121	12.8212	12.6644
8	47.3044	12.9018	12.6967
9	47.3559	12.9137	12.6529
10	47.2555	12.9107	12.6268
11	47.2380	13.1401	12.6538
12	47.3041	13.0615	12.5971

In all these cases the means are about the same as the correct mean, showing that a correction is not needed for the mean when data are grouped. All the means are nearly equal to the true mean, 47.3060 pedicels. Sometimes these means are a little larger than the true mean and sometimes a little smaller. The mean of the distribution for groups of 3 is 47.3104 pedicels, and the mean for the distribution for groups of 8 is 47.3044 pedicels. For this example it is not necessary to make adjustments for the arithmetic average when the data are grouped. It is shown in more advanced work in statistics that it is not necessary to make corrections on the arithmetic mean when the items are grouped into classes.*

This is not true of standard deviations. The unadjusted standard deviations get larger as classes are made larger, which is seen for the example under consideration by examining the third column in the above table. The unadjusted σ arising from groupings of 3 is smaller than that obtained from groupings of 8 and 9, etc. To bring the unadjusted standard deviations nearer to the true value, adjustments are necessary. These adjusted or corrected σ's are found in the fourth column of the last table and were determined by using formula (3.17). In all cases adjusted σ's are about equal

* H. C. Carver, "Editorial," *The Annals of Mathematics*, Vol. I, 1930, page 111.

to the correct $\sigma = 12.6711$ pedicels, which is the σ obtained before the data were grouped.

This illustration clearly shows, for this example, that it is necessary to make adjustments in order to reduce the standard deviation of the distribution of class marks of the grouped data. The mean and σ obtained from the distribution of class marks may never be equal to the true mean and standard deviation; they are considered to be approximately equal to them. The differences between the true mean and standard deviation and those obtained from the distribution of class marks, after σ has been adjusted, are very small; the values of M and σ obtained from the grouped data, after σ has been adjusted, are used as estimates of the true values. Formula (3.17) is generally used when distributions have high contact at both ends of the range, that is, distributions which rise gradually from the base line at both ends of the range and have one maximum. Details will now be given for the above distribution for groups of eights.

No. of Pedicels per Cluster Open Limits	No. of Pedicels per Cluster Closed Limits	Class Marks	Frequency	d	df	d^2f
12–19	11 5–19 5	15 5	88	−5	−440	2,200
20–27	19.5–27.5	23.5	480	−4		
28–35	27.5–35.5	31 5	1,829	−3		
36–43	35.5–43.5	39.5	2,865	−2		
44–51	43.5–51.5	47.5	2,936	−1		
52–59	51 5–59 5	55.5	2,492	0		
60–67	59.5–67.5	63 5	1,295	+1		
68–75	67.5–75.5	71.5	574	2		
76–83	75.5–83.5	79.5	202	3		
84–91	83.5–91.5	87.5	44	4		
92–99	91.5–99 5	95.5	15	5		
			12,800		−13,113	46,725

$$M_d = -\ 13{,}113/12{,}800 = 1.02445^{-} \text{ class units.}$$

$$M_v = 55.5 + (-1.02445)8 = 47.3044 \text{ pedicels.}$$

The unadjusted standard deviation is

$$\sigma = (\sqrt{\mu'_{2\,.\,d} - M_d^2})8 = (\sqrt{3.65039 - 1.04950})\cdot 8$$

$$= 12.9018 \text{ pedicels.}$$

The adjusted standard deviation is

$$\sigma_{\text{adj.}} = \left(\sqrt{\mu'_{2\,:\,d} - M_d^2 - \frac{1 - \dfrac{1}{k^2}}{12}}\right)\cdot w$$

$$= \left(\sqrt{3.65039 - 1.04950 - \frac{1 - \frac{1}{64}}{12}}\right)8$$

$$= \sqrt{2.51886}\cdot 8 = 12.6967 \text{ pedicels.}$$

The correction is made on the quantity under the square-root sign.

Continuous Variates

When distributions are composed of continuous variates, that is, variates which may differ by infinitesimal amounts, quantity k which represents the number of variates which can be put into a class is infinite, as there can be an infinite number of values between the upper and lower class limits. For example, if weights of men are grouped into classes of 10 pounds, the class 89.5–99.5 might have any weight between these two values, and surely this is infinite in number. In the case of continuous variates formula (3.17) becomes

$$(3.18) \qquad \sigma_{\text{adj.}} = (\sqrt{\mu'_{2\,:\,d} - M_d^2 - \tfrac{1}{12}})\cdot w,$$

since now $1/k^2 = 0$.

The following example will illustrate steps in finding σ for a grouped set of continuous variates.

WEIGHTS OF FRESHMEN AT THE UNIVERSITY OF MICHIGAN.
MEASUREMENTS MADE TO THE NEAREST POUND

CLASSES	CLASS MARK	FREQUENCY	d	$df(d)$	d^2f	$(d+1)^2f$
89.5– 99 5	94.5	1	−5			
99.5–109 5	104 5	19	−4			
109.5–119 5	114 5	80	−3			
119.5–129.5	124 5	179	−2			
129.5–139.5	134 5	260	−1			
139.5–149.5	144.5	210	0			
149.5–159.5	154.5	140	+1			
159.5–169.5	164.5	62	2			
169 5–179.5	174.5	28	3			
179.5–189.5	184.5	11	4			
189.5–199.5	194.5	4	5			
199.5–209.5	204.5	4	6			
209.5–219.5	214.5	2	7			
		1,000		−489	3,183	2,205

$M_d = -0.489$ class unit.

$M_v = 144.5$ lb. $+ (-0.489)10$ lb. $= 139\ 61$ lb.

By formula (3.18) the adjusted standard deviation is

$$\sigma_{\text{adj.}} = \sqrt{3.183 - 0.23912 - 0.08333} \cdot 10 = (\sqrt{2.86055^-})10$$
$$= 16.9132 \text{ lb.}$$

PROBLEMS

1. Find the number of items within 1, 2, and 3 σ's of the mean of the following distribution of ray flowers per head of oxeye daisies taken from a plot near Ann Arbor, Michigan. Plot the data, and show the frequencies in a histogram.

NO. OF RAY FLOWERS PER HEAD	FREQUENCY	NO. OF RAY FLOWERS PER HEAD	FREQUENCY
11–12	3	21–22	391
13–14	428	23–24	34
15–16	722	25–26	9
17–18	741	27–28	4
19–20	667	29–30	1
			3,000

2. Construct a histogram of the following distribution of the pulse beats per minute in English convicts. (See *Biometrika*, Vol. II, page 21.)

Pulse Beats per Minute	No. of Convicts	Pulse Beats per Minute	No. of Convicts
45–48	2	81– 84	86
49–52	5	85– 88	62
53–56	17	89– 92	42
57–60	57	93– 96	15
61–64	90	97–100	18
65–68	150	101–104	9
69–72	120	105–108	5
73–76	131	109–112	3
77–80	109	113–117	3
			924

On the graph show the percentage of items within 1, 2, and 3 standard deviations of the mean.

3. Make the classes twice as large in the above problem, and compare the mean and standard deviation with those found before.

4. The next table contains lung capacities of women students at the University of Michigan. Capacities given to the nearest cubic inch.

No. of Cubic Inches	Frequency	No. of Cubic Inches	Frequency
79 5– 99.5	2	179.5–199.5	114
99.5–119 5	14	199.5–219 5	53
119.5–139 5	78	219.5–239.5	30
139.5–159.5	122	239.5–259.5	9
159.5–179.5	158		580

Find the percentage of items within 1, 2, 3, and 4 standard deviations of the mean. If a student is picked at random from this group, what is the probability that her lung capacity will be between 1 and 2 σ's to the right of the mean?

5. If M_v and M_w are the means of 2 sets of positive variates, prove that the mean of the items in both sets lies between the two means if $M_v \neq M_w$.

THE STANDARD DEVIATION OF THE COMBINATION OF SETS

The mean and standard deviation of 100 test scores are respectively $M_v = 74.3$ and $\sigma_v = 4$; the mean and standard deviation of

80 other scores on the same test are respectively $M_w = 75.2$ and $\sigma_w = 3.6$. Let it be required to find the mean and standard deviation of the 180 scores. The mean of the entire set of scores according to formula (2.3) is

$$M_{v+w} = \frac{100(74.3) + 80(75.2)}{180} = 74.70.$$

To find the standard deviation one must use a formula similar to (3.8) for the combined sets of scores. One must find the following

$$(3.19) \qquad \sigma_{v+w} = \sqrt{\frac{\text{Sum of squares of all scores}}{180} - (M_{v+w})^2}.$$

It is necessary to find the sum of the squares of all the scores. According to formula (3.8)

$$\sigma_v^2 = \frac{\Sigma v^2}{n} - M_v^2, \quad \text{or} \quad \frac{\Sigma v^2}{n} = \sigma_v^2 + M_v^2, \quad \text{or}$$

$$(3.20) \qquad \Sigma v^2 = n(\sigma_v^2 + M_v^2).$$

In a similar manner the sum of the square of the scores of the w variates is

$$(3.21) \qquad \Sigma w^2 = m(\sigma_w^2 + M_w^2),$$

where m is the number of variates. According to the last two formulas the sum of the squares for the first set of 100 scores is $\Sigma v^2 = 100(4^2 + 74.3^2)$, and the sum of the squares of the 80 test scores is $\Sigma w^2 = 80(3.6^2 + 75.2^2)$. Substituting these in formula (3.19) gives

$$\sigma_{v+w} = \sqrt{\frac{100(4^2 + 84.3^2) + 80(3.6^2 + 75.2^2)}{180} - (74.7)^2}$$

$$= \sqrt{\frac{1,007,089}{180} - 5,580.09} = \sqrt{14.85} = 3.854.$$

Formulas (3.20) and (3.21) enable one to obtain the general formula for determining the standard deviation for the combination of two sets of data in terms of means, standard deviations,

and numbers of items in the two sets. The formula for the standard deviation of the items in both sets is

$$(3.22)\qquad \sigma_{v+w} = \sqrt{\frac{n_1(\sigma_v^2 + M_v^2) + n_2(\sigma_w^2 + M_w^2)}{n_1 + n_2} - (M_{v+w})^2},$$

where n_1 and n_2 are the number of items in the sets respectively.

This formula can be extended to the standard deviation for the combination of the items in several sets. This formula is

$$(3.23)\qquad \sigma_{v_1+v_2+\ldots+v_s} =$$

$$\sqrt{\frac{n_1(\sigma_{v_1}^2 + M_{v_1}^2) + n_2(\sigma_{v_2}^2 + M_{v_2}^2) + \ldots + n_s(\sigma_{v_s}^2 + M_{v_s}^2)}{n_1 + n_2 + \ldots + n_s} - (M_{v_1+v_2+\ldots v_s})^2},$$

where n_i represents the number of items in the ith set of variates, M_{v_i} and σ_{v_i} are respectively the mean and standard deviation of the ith set, and $M_{v_1+v_2+\ldots+v_s}$ is the mean of the items in all sets.

Suppose that the mean and standard deviation of 20 items are respectively 21 inches and 2 inches, and that the 20th item is 24 inches. Let it be required to find the mean and standard deviation after the 20th item has been omitted. The mean of the 19 items is

$$M_{19} = \frac{20(21) - 24}{19} = 396/19 = 20.84 \text{ in.}$$

The following formula for the standard deviation of a difference can be derived like (3.22).

$$(3.24)\qquad \sigma_{v-w} = \sqrt{\frac{n_1(\sigma_v^2 + M_v^2) - n_2(\sigma_w^2 + M_w^2)}{n_1 - n_2} - (M_{v-w})^2}.$$

According to (3.24) the standard deviation of the 19 items is

$$\sigma_{19} = \sqrt{\frac{20(2^2 + 21^2) - 1(0^2 + 24^2)}{19} - (20.84)^2}$$

$$= \sqrt{438.11 - 434.31} = \sqrt{3.80} = 1.95^- \text{ in.}$$

since the standard deviation of one item is zero.

Formula (3.24) is very important, for many times it becomes necessary to eliminate items after the mean and standard deviation have been found. Observations which differ from the mean by more than 4 standard deviations are rare and do not really belong to the set. Sometimes items which are more than 3 σ's from the

mean are omitted. If for any reason variates are found to be inconsistent with the other items in the set they are omitted from the data. Many times one does not know they are inconsistent until compared with the M_v and σ_v. The following example illustrates this point.

Thirteen students were requested to measure the length of a blackboard 10 times and to find the mean and standard deviation of their measurements. The 13 means and σ's were used to find the mean and σ of the 130 measurements. After this was found, the mean of one student's measurements was 9 standard deviations from the mean of all measurements. After his measurements were omitted his mean was off more than 9 σ's. On talking to the student it was discovered that he had measured the top of the board 5 times and the bottom of the board 5 times. It was discovered that the top of the board was $\frac{3}{8}$ inches longer than the bottom, which was measured by the other students. Formula (3.23) was used for finding the σ for the combined set of items; formula (3.24) was used to eliminate the inconsistent measurements of the student who measured the top and bottom of the blackboard.

PROBLEMS

1. The means and standard deviations of heights of men in five dormitories are as follows:

Dormitory	Mean	S. D.	No. of Men
A	67.6	2.6	65
B	67.3	2.4	47
C	68.1	2.7	34
D	66 8	2.5	149
E	65.4	3.2	178
			473

Find the mean and standard deviation of the heights of all men in the five dormitories.

2. In problem 1, Dormitory E has 27 Japanese students. The mean and standard deviation of these heights are respectively 64.1 inches and 2.3 inches. Find the mean and standard deviation of the heights in dormitory E after heights of the Japanese have been omitted.

3. If σ_v and σ_w are the standard deviations of two sets and $\sigma_v > \sigma_w$, prove that $\sigma_{v+w} > \sigma_w$.

4. The following data give means and standard deviations of scores made on a certain test for seventh-grade students in wards of a certain city.

Ward	Mean	S. D.	No. of Students
1	69 8	4.2	42
2	70.4	4 7	65
3	73.4	4.4	24
4	71.6	3.9	54
5	74 3	4.0	93
6	72.0	5.1	17

These results were turned in at the superintendent's office. Find the mean and standard deviation for the entire city. Discuss the records for each ward in comparison with results for the city.

5. The following data pertain to weights of freshmen in three large universities.

University	Mean	S. D.	No. of Freshmen
1	138.6 lb.	16.7 lb.	907
2	139.2 "	16.4 "	1,102
3	138.9 "	17.3 "	841
			2,850

Find the average weight of freshmen and the standard deviation of all these weights.

6. The mean and standard deviation of 83 measurements are respectively 17.3 gallons and 1.3 gallons. It is found that one of the measurements, 22.3 gallons, was incorrectly made. Omit this item and find M and σ for the others.

QUARTILES

Quartiles of a distribution of variates are quantities which divide the distribution into 4 equal parts. The median is sometimes spoken of as the second quartile, Q_2. The first or lower quartile of a grouped frequency distribution is found by a formula similar to that for the median. This formula for the first or lower quartile is

$$Q_1 = C_1 + \left(\frac{\frac{n}{4} - n_1}{f_1}\right) w, \tag{3.25}$$

where C_1 is the lower limit of the class in which the $(n/4)$th item falls when the items are arranged in order; n_1 is the sum of the frequencies of the classes below C_1; f_1 is the frequency of the class in which the $(n/4)$th item falls; n is the sum of the frequencies in all classes; and w is the width of the class in which the $(n/4)$th item falls.

The third quartile is found by the formula

$$Q_3 = C_3 + \left(\frac{\frac{3n}{4} - n_3}{f_3}\right) w, \tag{3.26}$$

where C_3, f_3, and w are respectively the lower limit, the frequency and the width of the class in which the $(\frac{3}{4}n)$th item falls, n_3 is the sum of all frequencies below this class, and n is the total frequency.

One-half of the items of a frequency distribution falls between the first and third quartiles.

Another measure of dispersion is

$$Q = \frac{Q_3 - Q_1}{2}, \tag{3.27}$$

and is called the quartile deviation or the semi-interquartile range.

The following example will illustrate the steps in finding the semi-interquartile range.

Lung Capacity of Women	No. of Women
79.5– 99.5	2
99.5–119.5	14
119.5–139.5	78
139.5–159.5	122
159 5–179.5	158
179.5–199.5	114
199 5–219.5	53
219.5–239.5	30
239.5–259.5	9
	580

$$Q_1 = 139.5 + \left(\frac{580/4 - 94}{122}\right)20 = 139.5 + 8.36 = 147.86 \text{ cu. in.}$$

$$Q_3 = 179.5 + \left(\frac{3 \cdot 580/4 - 374}{114}\right)20 = 179.5 + 10.70 = 190.20 \text{ cu. in.}$$

from which

$$Q = \frac{Q_3 - Q_1}{2} = \frac{190.20 - 147.86}{2} = 21.17 \text{ cu. in.}$$

Fifty per cent of the distribution lies between Q_3 and Q_1, and *about* 50 per cent lies within a distance of Q from the median. For symmetrical distributions there is exactly 25 per cent of the distribution within a distance of Q to the right of the median and 25 per cent within a distance of Q to the left. The above distribution is exhibited in Fig. 3.3 by a histogram, together with the quartiles.

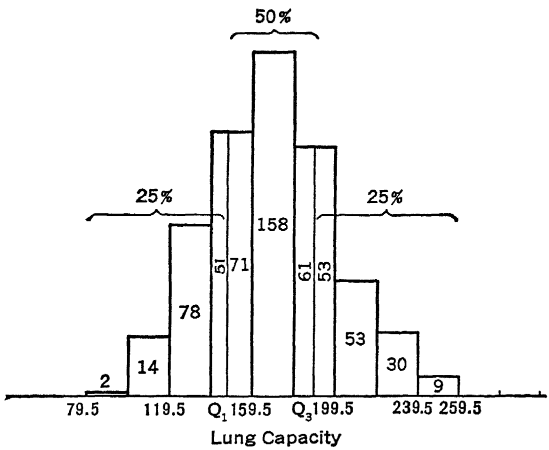


FIG. 3.3.—Distribution of lung capacities of women students, showing the quartiles.

THE COEFFICIENT OF VARIATION

The coefficient of variation is equal to the ratio of the standard deviation to the mean multiplied by 100 and is

$$V = \frac{100\sigma_v}{M_v}. \tag{3.28}$$

This is an abstract number and shows the variability of the items in the distribution. It is a very good quantity for comparing two distributions. If the mean and standard deviations of one set of variates are respectively 37 inches and 2 inches, and the mean and standard deviations of another set of variates are respectively

37 inches and 3.2 inches, the coefficient of variation enables one to compare the variabilities of the two sets. These coefficients are

$$V_1 = 100(2/37) = 5.4 \quad \text{and} \quad V_2 = 8.6,$$

which show that there is much more variability in the second set than in the first.*

If standard deviations of two distributions are equal and the means are not the same, the coefficients of variation may not give any information concerning the variabilities of the distributions. Let both σ's be equal to 10 gallons. $M_1 = 40$ gallons, and $M_2 = 60$ gallons. The coefficients of variation are, respectively, $V_1 = 25$, $V_2 = 17$. These values would indicate that the first was more variable or showed more variability than the second. There is no difference between the amount of variability of the two distributions. The main use of V is for comparison of distributions which have means which are about the same and for comparisons of distribution which are expressed in different units. Heights of girls 6 years of age and of girls 19 years furnish an example of this point. The mean and standard deviations of heights of 6-year-old girls are respectively 45.7 inches and 2 inches, while these for 19-year-old girls are respectively 62 inches and 2 inches. One cannot say that the heights of the 6-year-old girls are more variable than the heights of 19-year-old girls because V is greater in the former. The percentages of items within 1, 2, and 3 σ's of the mean indicate how the items are scattered about the means and would show variability much better than V, for this situation. The quantity V is useful, but care should be taken when interpreting results.

The coefficient of variation is the percentage the standard deviation is of the mean.

PROBLEMS

1. If the first quartile is 142 and the semi-interquartile is 18, what is the third quartile?

2. The coefficient of variation is 40 and the mean is 30; find σ.

3. The following frequency distributions pertain to weights of light and heavy men. Find the coefficients of variation for the two distributions and discuss the results.

* For test of significance see chapter on Standard Errors of Statistics.

Weights of Men	No. of Men	Weights of Men	No. of Men
89.5– 99 5	2	199.5–209.5	8
99.5–109 5	11	209.5–219 5	30
109.5–119.5	36	219.5–229.5	57
119 5–129.5	71	229.5–239 5	69
129 5–139.5	96	239.5–249.5	74
139 5–149.5	49	249.5–259.5	39
149 5–159.5	21	259 5–269 5	17
159.5–169.5	9	269 5–279.5	8
169.5–179.5	3	279.5–289.5	3
179.5–189.5	1	289 5–299.5	2
	299	299.5–309.5	1
			308

4. Find the semi-interquartile range for the scores made at a certain university by students in the mathematics department.

Scores	No. of Students	Scores	No. of Students
Under 50	55	75–79	129
50–54	31	80–84	95
55–59	49	85–89	78
60–64	73	90–94	20
65–69	102	95 and above	9
70–74	144		785

How many items are within 1 semi-interquartile range of the median?

5. The following distributions of ligulate flowers of oxeye daisies were taken respectively at the beginning and end of the flowering season. Compare the coefficient of variation for the two samples.

No. of Ligulate Flowers per Head First of Season	No. of Heads	No. of Ligulate Flowers per Head Last of Season	No. of Heads
11–12	0	11–12	3
13–14	131	13–14	139
15–16	242	15–16	241
17–18	238	17–18	240
19–20	229	19–20	232
21–22	146	21–22	128
23–24	10	23–24	17
25–26	3	25–26	0
27–28	0	27–28	0
29–30	1	29–30	0
	1,000		1,000

6. Use an ogive to find the quartiles of the first distribution in problem No. 3.

7. Set up formulas for finding the percentiles of a frequency distribution.

8. Use these formulas for finding the percentiles of the distribution in No. 4.

9. Set up formulas for finding the deciles of a frequency distribution.

CHAPTER 4

THE NORMAL CURVE

PROPERTIES OF THE NORMAL CURVE

One of the most important frequency distributions in statistics is the normal distribution. Since it plays such an important role it is essential to know some of its properties. The equation of the normal curve is

$$(4.1) \qquad Y = \frac{1}{\sqrt{2\pi}\,\sigma_v} e^{-\frac{(v - Mv)^2}{2\sigma_v^2}},$$

where $e = 2.718$ and $\pi = 3.1416$, these being approximate values of π and e. If $M_v = 0$ and $\sigma_v = 1$, the above equation becomes

$$(4.2) \qquad Y = \frac{\quad}{\sqrt{2\pi}} e^{-v^2/2} = 0.3989\, e^{-v^2/2}.$$

Several tables have been constructed for values of e^{+s} and e^{-s} for various values of s. The table below contains values of $e^{-\frac{v^2}{2}}$ and of Y for various values of v. Values of Y are plotted in Fig. 4.1.

v	$e^{-v^2/2}$	Y	v	$e^{-v^2/2}$	Y
−3 0	0.0111	0.0044	+0.2	0 9802	0.3910
−2.5	0.2437	0.0175	+0 5	0.8781	0.3521
−2.0	0.1353	0.0540	+1.0	0.6065	0.2420
−1.5	0.3230	0.1295	+1.5	0.3230	0 1295
−1.0	0.6065	0.2420	+2.0	0.1353	0.0540
−0.5	0 8781	0.3521	+2.5	0.0437	0 0175
−0.2	0.9802	0.3910	+3.0	0.0111	0.0044

On examining the equation of the curve in Fig. 4.1 it is seen that the curve approaches the horizontal axis very rapidly to the left of −3 and to the right of +3, that the curve is symmetrical and has one maximum, which is at the origin. The area under the curve above the horizontal axis is unity; the area within 1 unit of the origin is about 68 per cent of the total area, or the area within 1 standard deviation ($\sigma = 1$ here) of the origin or the mean

is about 68 per cent of the entire area. About 95 per cent of the area is within 2 units or standard deviations of the origin; about 99 per cent of the area is within 3 σ's of the origin and about 99.999 per cent of the area is within 4 units or 4 σ's of the origin, which is the mean of the distribution. The curve has inflection points at +1 and −1. One half of the area is within 0.6745 unit of the origin or the mean. The curve decreases very rapidly to

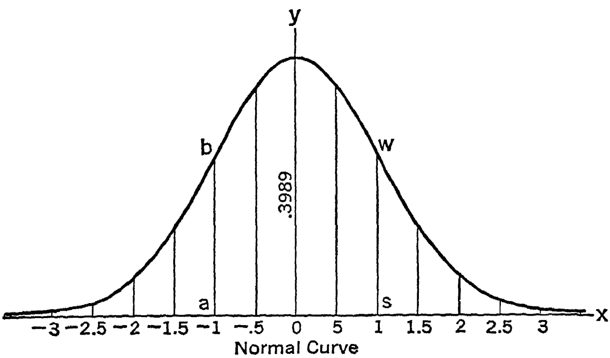


FIG. 4.1.

the right of +0.5 and to the left of −0.5. The slope of the curve to the left of −3 and to the right of +3 is nearly zero.

The ogive of this normal curve is plotted in Fig. 4.2 and shows the amount of area under the curve in Fig. 4.1 from $-\infty$ to various values of v. The ogive starts at zero at $-\infty$ and rises very slowly until $v = -2$ is reached, where it rises more rapidly until $v = +2$ is reached; beyond +2 the ogive gradually approaches the line $Y = 1$. At $+\infty$ it is equal to +1. At the origin the height of the ogive above the horizontal axis is +0.5, as it should be, since the area under the curve in Fig. 4.1 from $-\infty$ to 0 is one-half of the entire area under the curve. The line AB in Fig. 4.2 gives the area under the curve in Fig. 4.1 to the left of ab, while the length of the line SW in Fig. 4.2 gives the area under the curve in Fig. 4.1 to the left of sw.

Table I gives values of the ogive curve for the normal curve for various values of t, where the equation of the normal curve is $\frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$. Table II gives ordinates of the normal curve for various values of t.

Let it be required to find the area under the normal curve between 0.3 and 0.9. From the ogive curve in Table I the area to the left of 0.3 under the normal curve is 0.6179, while the area to the left of 0.9 and under the normal curve is 0.8159; hence the area between 0.3 and 0.9 is $0.8159 - 0.6179 = 0.1980$. The area between -0.67 and $+0.67$ is equal to $0.7486 - 0.2514 = 0.4972$, which is nearly one-half of the area under the curve Y.

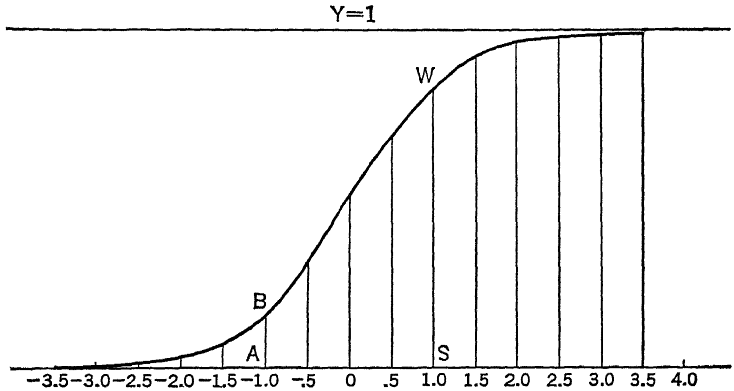


FIG. 4.2.—Ogive of the normal curve.

Suppose that a distribution of variates is normal with $\sigma = 1$ and that there are 20,000 variates in the distribution. Let it be required to find the number between 0.4 and 0.8 unit to the right of the mean of the distribution. Imagine that the origin of a normal distribution coincides with the mean of the given distribution. The area between 0.4 and 0.8 from Table I is $0.7881 - 0.6554 = 0.1327$, which is 13.27% of the entire area; hence the number of variates of the 20,000 between these limits is $20,000(0.1327) = 2,654$. The number between -2 and -1.6 is equal to $20,000(0.03205) = 641$. Since areas under the normal curve from $-\infty$ to t are given in percentages it is easy to find the number of variates in any normal distribution between two limits or within any class.

Many teachers believe that grades for a large number of students are normally distributed. Assuming this to be true, let it be required to find the percentages of A's, B's, C's, D's, and E's. Assume that the normal curve extends for practical purposes from -3 to $+3$, or extends over a range of 6 units. Since there are

5 grade groups each will cover 6/5 units on the horizontal axis. The values of these limits for the grade groups in terms of the t's are

$$-3, -1.8, -0.6, +0.6, +1.8, +3.$$

By utilizing Table I the following percentages are found for these grades:

Percentage of A's = 3.5, percentage of B's = 23.8,
" " C's = 45.1, " " D's = 23.8,
" " E's = 3.5, " beyond -3 and $+3$ = 0.3.

These are plotted in Fig. 4.3.

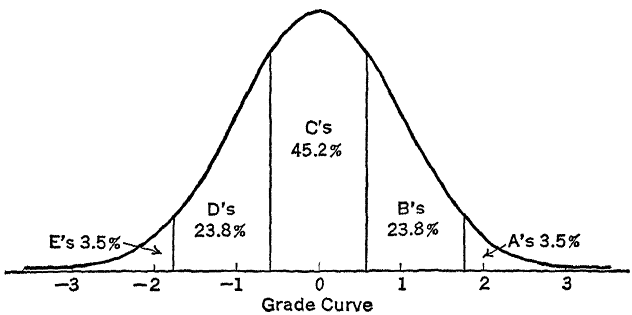


FIG. 4.3.—Percentages of grades on the assumption that the grade curve is normal.

GRADUATION

In general there are very few distributions with unity as standard deviation. To be able to use Tables I and II it is necessary to change the original units into standard units or to change v's into t's. Tables I and II are in terms of the standard unit t. To change from any unit to standard units it is necessary to use t equal to

$$t_i = \frac{v_i - M_v}{\sigma_v} = \frac{\bar{v}_i}{\sigma_v}. \tag{4.3}$$

The mean of the t's is zero, as can be seen by finding the average of the t's from (4.3); the standard deviation of the t's is unity, as can be seen by finding the average of the squares of the t's. The normal curve in terms of t is

$$\frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(t-M_t)^2}{2\sigma_t^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}},$$

since $M_t = 0$ and $\sigma_t = 1$. The following illustration will show how a distribution can be compared with a normal distribution, or how a distribution can be graduated by a normal curve. The first two columns in the following table contain the distribution which is to be graduated. To be able to compare frequencies of this set of variates with those of the normal, it is necessary to change the units from inches to standard units by use of formula (4.3).

Heights of Adult Males Born in England, Ireland, Scotland and Wales

(The Anthropometric Committee to the British Association, Report 18,893, p. 256. Original measurements made to the nearest $\frac{1}{8}$ inch.)

Classes $56\frac{15}{16}$ to $57\frac{15}{16}$, $57\frac{15}{16}$ to $58\frac{15}{16}$, etc.

Heights	Frequency	t	Area Below t	Difference of Areas	Expected Frequencies
$57\frac{7}{16}$	2	−4.09	0.00002	0.00009	1
$58\frac{7}{16}$	4	−3.70	.00011	.00036	3
$59\frac{7}{16}$	14	−3.31	.00047	.00128	11
$60\frac{7}{16}$	41	−2.92	.00175	.00379	33
$61\frac{7}{16}$	83	−2.54	.00554	.01024	88
$62\frac{7}{16}$	169	−2.15	.01578	.02342	201
$63\frac{7}{16}$	394	−1.76	.03920	.04614	396
$64\frac{7}{16}$	669	−1.37	.08534	.07820	671
$65\frac{7}{16}$	990	− .98	.16354	.11405	979
$66\frac{7}{16}$	1,223	− .59	.27759	.14315	1,229
$67\frac{7}{16}$	1,329	− .20	.42074	.15460	1,327
$68\frac{7}{16}$	1,230	+ .19	.57534	.14360	1,234
$69\frac{7}{16}$	1,063	+ .58	.71904	.11243	965
$70\frac{7}{16}$	646	+ .96	.83147	.08002	687
$71\frac{7}{16}$	392	+1.35	.91149	.04758	409
$72\frac{7}{16}$	202	+1.74	.95907	.02434	209
$73\frac{7}{16}$	79	2.13	.98341	.01072	92
$74\frac{7}{16}$	32	2.52	.99413	.00406	35
$75\frac{7}{16}$	16	2.91	.99819	.00132	11
$76\frac{7}{16}$	5	3.30	.99951	.00038	3
$77\frac{7}{16}$	2	3.69	.99989	.00009	1
$78\frac{7}{16}$	0	4.07	.99998	.000016	0
$79\frac{7}{16}$	0	4.46	.999996		
	8,585				8,585

$M_v = 67.4584$ in.; $\sigma_v = 2.5723$ in.

The t's which correspond to the v's were found in the above table by formula (4.3) after the mean and standard deviation were obtained. The t corresponding to the lower limit of the first class is

$$t_1 = \frac{v_1 - M_v}{\sigma_v} = \frac{56.9375 - 67.4584}{2.5723} = \frac{1}{2.5723}(-10.5209) = -4.09.$$

The value of t corresponding to the lower limit of the second class is

$$t_2 = \frac{57.9375 - 67.4584}{2.5723} = \frac{1}{2.5723}(-9.5209) = -3.70,$$

where the quantity in the last parentheses is greater than the quantity in the parentheses above by 1 unit. If one has access to a computing machine the t's can be obtained by turning the crank handle of the machine backward once for each t, since the quantities in the parentheses for successive t's differ by unity. These t values are listed in column 3 in the above table.

The fourth column in the above table contains areas under this normal curve to the left of the various values of t; these values were taken from Table I. The first value in column 4 is the area under the normal curve to the left of $t = -4.09$; the next value is the area to the left of -3.70, etc. Quantities in column 5 are areas under this normal curve between successive values of t. The fourth quantity in this column is the area between $t = -2.54$ and $t = -2.92$, and is found by subtracting the two quantities 0.00175 and 0.00554 in column 4. Quantities in column 5 are found by subtracting successive quantities in column 4. Quantities in column 5 are the percentages of frequencies for classes of a distribution which is normal. Expected frequencies for classes of the distribution of 8,585 heights are found by multiplying the percentages for the various classes by 8,585. These frequencies, called expected frequencies, are found in the last column in the table. Values in the last column are sometimes called graduated values of the original frequencies. The sum of the expected frequencies is 8,585, which is the sum of all frequencies in the original distribution. The original distribution is not a normal distribution but approximates one.

The percentage of heights within 1 σ of the mean is found by adding the frequencies between -1 and $+1$ and is about 68 per cent of the total distribution.

PROBLEMS

1. Graduate the following distribution by the normal distribution,* if the mean = 7.0055 and the standard deviation = 0.121.

Classes Closed Limits	Frequency	Classes Closed Limits	Frequency
6.625–6 675	0	7 025–7.075	55
6 675–6.725	2	7 075–7 125	50
6.725–6.775	12	7.125–7 175	40
6.775–6.825	20	7.175–7.225	15
6.825–6.875	22	7 225–7.275	11
6.875–6 925	46	7.275–7.325	3
6.925–6.975	57	7.325–7.375	1
6.975–7.025	66		400

2. By using percentages of grades on page 86 find the number of students making A's, B's, C's, D's, and E's respectively if there are 300 students.

3. Assume that the range of the normal curve for practical purposes is 7 instead of 6; find the percentages of A's, B's, C's, D's, and E's for the grade curve if it is normal.

4. Assume that the grade curve is normal and that each group is divided into three equal parts, C−, C, C+, B−, B, B+, etc. Find the percentages of the different grades. Assume that the range is from −3 to +3.

5. The following scores were made on an intelligence test. Assume that these scores are distributed normally; find the expected frequencies for the various score groups. What can be said about the test as measuring intelligence?

Scores	Frequency	Scores	Frequency
34.5–39.5	3	69 5–74.5	109
39.5–44.5	17	74.5–79.5	78
44.5–49.5	28	79.5–84.5	46
49.5–54.5	49	84.5–89.5	31
54.5–59.5	74	89.5–94.5	15
59.5–64.5	113	94.5–99.5	5
64.5–69.5	124		

6. Plot the normal curve from values given in Table II.

* "An Application of Thiele's Semi-Invariants to the Sampling Problem" by C. C. Craig, *Metron*, Vol. 7, No. 4, 1928, pages 3–75.

7. If $t_i = 2.4$, $M_v = 89.4$ feet, and $\sigma_v = 3.7$ feet, find the v that corresponds to this t.

8. Plot the ogive of the normal curve from values given in Table I.

9. If $t_i = -2.3$, $v_i = 200$ gallons, and $\sigma_v = 7.2$ gallons, find M_v.

10. If $t = -0.72$, $v = 34.6$ cc., $M_v = 37.8$ cc., find σ_v.

ORDINATES OF THE NORMAL CURVE

The histogram of the following normal distribution is plotted in Fig. 4.4.

TABLE 4.1

v	$f(v)$	v	$f(v)$	
0	1	8	175	$M_v = 7.0,$
1	2	9	121	$\sigma_v = 2.0.$
2	9	10	66	
3	28	11	28	
4	66	12	9	
5	121	13	2	
6	175	14	1	
7	196		1,000	

Let it be required to construct a histogram of this distribution so that the variable shall be

$$t_i = \frac{v_i - M_v}{\sigma_v},$$

instead of v_i, and so that the sum of the frequencies shall be unity instead of 1,000. This shall be done in two steps which shall be described in detail.

The t variates according to the above value of t run from $t_1 = (0 - 7)/2$ to $t_{15} = (14 - 7)/2$ for class marks.

Let it first be required to construct a histogram to have the same frequencies as the histogram of data in Table 4.1 but with t as the variable instead of v. From the definition of t it is seen that the bases of the rectangles of the required histogram will be equal to the bases of the histogram in Fig. 4.4 except they will be divided by $\sigma = 2$. Since the bases of rectangles in the required histogram are equal to $1/\sigma = \frac{1}{2}$ the altitudes of these rectangles must be equal to $\sigma f(v)$ so as to contain the same areas as the rectangles in Fig. 4.4, or so as to have the same corresponding frequencies. The new histogram possesses rectangles with bases

equal to one-half of the base of the rectangles in Fig. 4.4, and altitudes twice as high as those in Fig. 4.4. The area of rectangle $APRT$ in Fig. 4.4 is equal to $1 \times 66 = 66$, while the area of the corresponding rectangle $aprt$ in Fig. 4.5, which is the first required histogram, is $(1/2) \times 132 = 66$. Since the bases in

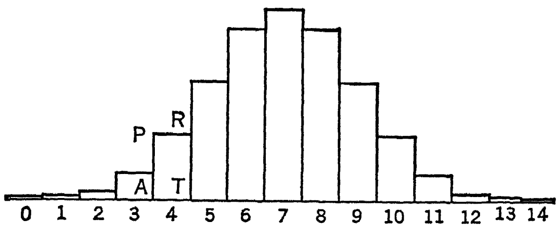


FIG. 4.4.—Histogram of the distribution in Table 4.1. The $f(v)$ distribution.

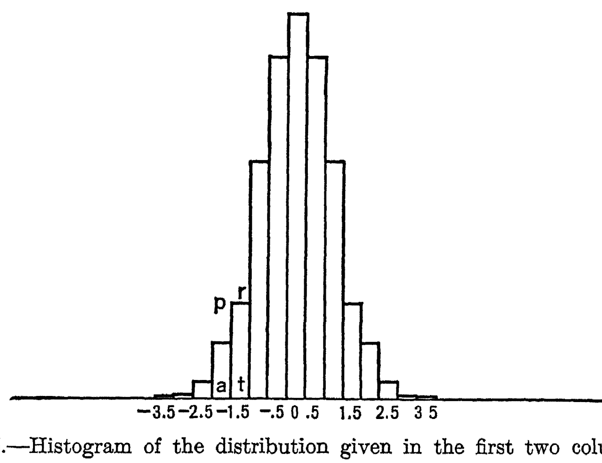


FIG. 4.5.—Histogram of the distribution given in the first two columns of Table 4.2. The F(t) distribution.

Fig. 4.5 are one-half of the size of the bases in Fig. 4.4, the altitude of the former must be twice as great as those in the latter, in order to have corresponding areas the same. Mid-points of rectangles of Fig. 4.5 are given in Table 4.2 in the t column. Altitudes are given in the column headed by $F(t) = \sigma_v \cdot f(v)$, where $F(t)$ represent the altitudes of the first required histogram. Frequencies of the $F(t)$ distribution or histogram are listed in column 3; they are equal to the respective areas of the rectangles.

TABLE 4.2

t	$F(t) = \sigma f(v)$	Area of Rectangles of $F(t)$ Curve $\frac{1}{\sigma}\cdot\sigma f(v)$	Area of Rectangles of $f_1(t)$ Curve $\frac{1}{\sigma}\cdot\frac{\sigma f(v)}{N}$	Ordinates or Altitudes of Rectangles of $f_1(t)$ Curve $\frac{\sigma f(v)}{N}$
−3.5	2	1	0 001	0.002
−3.0	4	2	.002	.004
−2.5	18	9	.009	.018
−2.0	56	28	028	.056
−1.5	132	66	.066	132
−1.0	242	121	.121	.242
−0.5	350	175	.175	.350
0.0	392	196	.196	.392
+0.5	350	175	.175	.350
+1.0	242	121	.121	.242
1.5	132	66	.066	.132
2.0	56	28	.028	.056
2.5	18	9	.009	.018
3 0	4	2	.002	.004
3.5	2	1	.001	.002

Let it now be required to construct a histogram which shall have bases of the same size as those in the histogram for $F(t) = \sigma f(v)$,

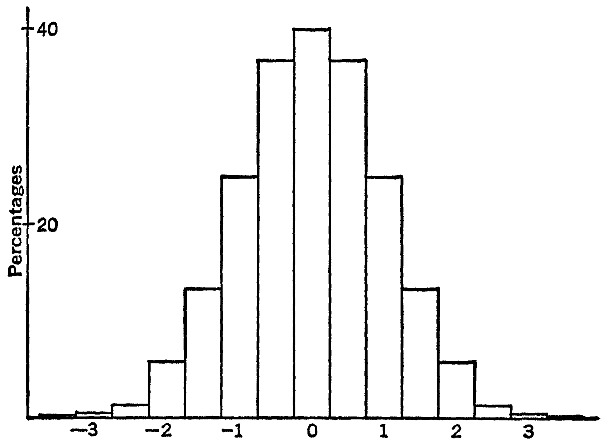


FIG 4.6.—Histogram of the distribution given in the first and last columns of Table 4.2. The $f_1(t)$ distribution.

as shown in Fig. 4.5, and with rectangles the sum of whose areas shall equal unity. To do this, each altitude of the rectangles in Fig. 4.5 must be divided by $N = 1{,}000$; this makes the altitudes of the rectangles in the second required histogram equal to $\dfrac{\sigma f(v)}{N}$. The second required histogram is shown in Fig. 4 6; the bases of the rectangles of this histogram are equal to $1/\sigma$, and the altitudes are equal to $f_1(t) = \dfrac{\sigma f(v)}{N}$, which gives area equal to $(1/\sigma)[\sigma f(v)/N] = f(v)/N$. The areas of the rectangles in Fig. 4 6 are the respective areas in Fig. 4.4 expressed as percentages. Now we have a histogram in Fig. 4.6 whose bases are equal to the bases of the original histogram in Fig. 4.4 divided by σ, and whose total frequency is unity. These changes are shown in the above table.

The above method of going from the $f(v)$ distribution to the $f_1(t)$ distribution was given to show the relation between $f(v)$ and $f_1(t)$, which shows the nature of Table II. Table II contains ordinates of the $f_1(t)$ curve. One can go directly from $f(v)$ to $f_1(t)$ by the formula

$$f_1(t) = \frac{\sigma}{N} f(v), \tag{4.4}$$

because

$$f_1(t) = \frac{F(t)}{N} = \frac{\sigma f(v)}{N}.$$

One can go directly from $f_1(t)$ to $f(v)$ by the formula

$$f(v) = \frac{N}{\sigma} f_1(t), \tag{4.5}$$

because

$$f(v) = \frac{1}{\sigma} F(t) = \frac{N}{\sigma} f_1(t).$$

Formula (4.5) is of great importance in graduating distributions by ordinates of the normal curve given in Table II. Of course, this is for curves which are approximately normal, or for comparing distributions with normal distributions. An example will be given to illustrate the method of graduating distributions by ordinates of the normal curve.

The function $f_1(t)$ is always known, for its ordinates are given

in Table II and they can be used to graduate any distribution since they are given in percentages. For any particular distribution the t's are found from the mean and standard deviation. From values of the t's one finds the ordinates of the desired curve, or the expected frequencies, on the assumption that it is normally distributed. From formula (4.5) the expected values of the v curve are obtained. These are the values of $f(v)$ on the assumption that it is normal.

The following table contains the distribution of the number of questions answered on an intelligence test by 10,000 students.

No. of Questions Answered = v	Frequency	t	$f(t)$	Expected Frequency
36	7	−3.50	0.00087	4
37	18	−3.00	.00443	22
38	80	−2.50	.01753	87
39	260	−2.01	.05292	264
40	640	−1.51	.12758	635
41	1,230	−1.01	.23955	1,193
42	1,740	−0.51	.35029	1,745
43	1,980	−0 01	.39892	1,987
44	1,760	+0.48	.35553	1,771
45	1,200	+0.98	.24681	1,229
46	680	1.48	.13344	665
47	280	1.98	.05618	280
48	94	2.48	.01842	92
49	22	2.98	.00471	23
50	9	3.47	.00097	5
	10,000			10,002*

$M_v = 43.0268$ questions, $\sigma_v = 2.0077$ questions.

$$t_1 = \frac{36 - 43.0268}{2.0077} = -3.50, \quad t_2 = \frac{37 - 43.0268}{2.0077} = -3.00, \text{ etc.}$$

The fourth column in the above table was found from Table II for respective values of t; these are the ordinates of the normal curve. The fifth column was found from formula (4.5). The first number in this column was found as follows:

$$f(v_1) = \frac{10,000}{2.0077}(0.00087) = 4980.82\,(0.00087) = 4 \text{ students.}$$

* Because of rounding off decimals.

The fifth number in the last column was found as follows:

$$f(v_5) = 4980.82(0.12758) = 635 \text{ students.}$$

The other expected frequencies were found similarly.

PROBLEMS

1. Graduate the following distribution by use of ordinates of the normal curve.

Time of Running 1 Mile for 1 Track Man	No. of Days	Time of Running 1 Mile for 1 Track Man	No. of Days
4 min. 18 sec.	1	4 min. 24 sec.	17
4 " 19 "	3	4 " 25 "	13
4 " 20 "	7	4 " 26 "	6
4 " 21 "	12	4 " 27 "	2
4 " 22 "	18	4 " 28 "	1
4 " 23 "	20		

2. A distribution is normal with mean = 140.4 pounds and standard deviation is 16.6 pounds. If there are 347 items between 154 pounds and 162 pounds, including 154 and 162, find the number of items, N, in the entire distribution. Original measurements made to the nearest $\frac{1}{2}$ pound.

3. In a normal distribution with mean = 37.4 inches and $\sigma = 1.8$ inches the number of items with measurement 35 inches is 509. Find the total number of items in the entire distribution.

4. In a normal distribution with mean = 16 grams and $\sigma = 1.3$ grams, the number of items between 13.6 and 14.2, including 13.6 and 14.2, together with those between 17.3 and 18.6, including 17.3 and 18.6, is 804. Find the number of items in the whole distribution if the original measurements were made to the nearest 0.1 gram.

5. In a distribution with $M_v = 99.7$ feet and $\sigma_v = 5.4$ feet, there are 478 items from 106 to some point h not including h. If the distribution is normal, find the value of h, provided there are 2,468 items in the entire distribution, and the original measurements are made to the nearest unit.

6. A drygoods store finds that the sale of ladies' hose approximates a normal distribution. Find the number of each size it should have on hand to supply 20,000 ladies if sizes range from 7, $7\frac{1}{2}$, 8, $8\frac{1}{2}$ . . ., 11. Assume that the range of the normal curve is from -3.6 to $+3.6$.

CHAPTER 5

SKEWNESS AND KURTOSIS

SKEWNESS

In an earlier chapter it was shown that the sum of the deviations of the variates from the mean was equal to zero, the average of the absolute values of these deviations was equal to the mean deviation, and the square root of the average of the squares of these deviations was defined as the standard deviation. For normal distributions and approximately normal distributions there is about 68 per cent of the variates within 1 σ of the mean, about 95 per cent within 2 σ's of the mean, and about 99.9 per cent within 3 σ's. Very few variates are beyond 3 standard deviations of the mean. It was also pointed out that the inflection points of the normal curve were located 1 σ to the right of the mean and 1 σ to the left. The square root of the average of the squares of the deviations from the mean proved to be a very important characteristic of frequency distributions; it will prove of more importance as a measure of dispersion in material which follows.

This chapter presents the importance of the averages of the cubes of the deviations of the variates from the mean and also the value of the average of the fourth powers of these deviations.

Let us consider again the normal distribution given on page 90, which is listed in the next table and graphically shown in Fig. 4.4.

Deviations of the variates from the real mean are listed in the third column of the above table. Column 4 contains the cubes of these deviations multiplied by their respective frequencies. Since the distribution is symmetrical with respect to the mean, the cubes of the negative deviations are equal to the corresponding cubes of the positive deviations; hence the average of the cubes of these deviations is zero. This is always true of symmetrical frequency distributions and, of course, of normal distributions.

Figure 5.0 presents graphically the above normal distribution. It has one maximum point and contains as much area to the left

v	$f(v)$	$\bar{v}=v_i-M_v$	$\bar{v}^3\cdot f(v)$	$\bar{v}^4\cdot f(v)$
0	1	−7	− 343	2,401
1	2	−6	− 432	2,592
2	9	−5	−1,125	5,625
3	28	−4	−1,792	7,168
4	66	−3	−1,782	5,346
5	121	−2	− 968	1,936
6	175	−1	− 175	175
7	196	0	−6,617	000
8	175	+1	+ 175	175
9	121	+2	+ 968	1,936
10	66	+3	+1,782	5,346
11	28	4	1,792	7,168
12	9	5	1,125	5,625
13	2	6	432	2,592
14	1	7	343	2,401
	1,000		+6,617	50,486
		Total	0,000	

of the mean as to the right. The curve rises gradually at 3 σ's to the left of the mean, rises rapidly between $-(3/2)\sigma$ and $-(1/2)\sigma$, and reaches its maximum at the origin, in the neighborhood of

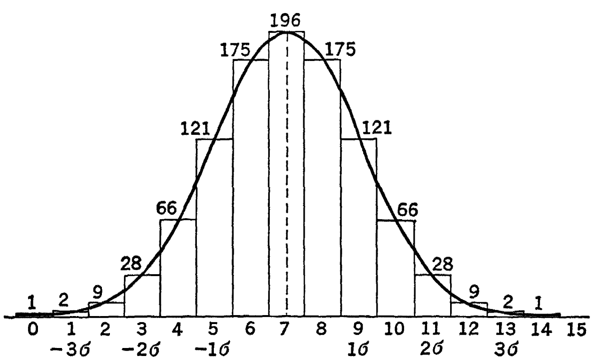


FIG. 5.0.—Histogram and frequency curve of a normal distribution.

which it has very little slope. To the right of the mean it does the same things but in reverse order, which makes it a symmetrical curve. It is a curve with zero skewness.

If the average of the cubes of the deviations of the variates from

the mean is not equal to zero the distribution is not normal or symmetrical; hence this average of the cubes indicates whether or not a frequency distribution is asymmetrical or skew. Various measures of skewness have been given by other authors. A measure for skewness, or for the amount of departure from symmetry, which is being adopted by a large number of statisticians, is the average of the cubes of the deviations of the variates from the mean divided by the cube of the standard deviation, or

$$\alpha_{3 \cdot v} = \frac{\mu_{3 \cdot v}}{\sigma_v{}^3}. \tag{5.1}$$

In other words, skewness is the third moment about the mean divided by the cube of the standard deviation. Dividing by σ^3 makes the skewness of a distribution an abstract quantity, which does not depend upon the unit of the original variates.

If the sum of the positive deviations of the variates from the mean is greater than the sum of the negative deviations the skewness is positive, and vice versa.

Table 5.1 contains a skew frequency distribution and computations necessary for finding skewness.

TABLE 5.1

NUMBER OF GALLONS OF GASOLINE USED ON A CERTAIN ROUTE BY A BUS

No. of Gallons v	$f(v)$	$\bar{v}$	$\bar{v}^2 f(v)$	$\bar{v}^3 f(v)$
35	5	−4	80	− 320
36	42	−3	378	−1,134
37	145	−2	580	−1,160
38	249	−1	249	− 249
39	196	0	000	−2,863
40	178	+1	178	+ 178
41	101	2	404	+ 808
42	51	3	459	1,377
43	19	4	304	1,216
44	9	5	225	1,125
45	4	6	144	864
46	1	7	49	343
	1,000		3,050	+5,911
				+3,048

$$\mu_v = 39 \text{ gallons,}$$

$$\mu_{2:v} = 3.05; \qquad \sigma_v = 1.746 \text{ gallons;}$$

$$\mu_{3:v} = 3.048; \qquad \alpha_{3:v} = \frac{\mu_{3 \cdot v}}{\sigma_v^3} = \frac{3.048}{5.325} = 0.5724.$$

The distribution in Table 5.1 is plotted in Fig. 5.1 with the plot of a normal distribution. Compare the part of the curve A to the right of the mean with the part to the left. Notice that the maximum is to the left of the mean.

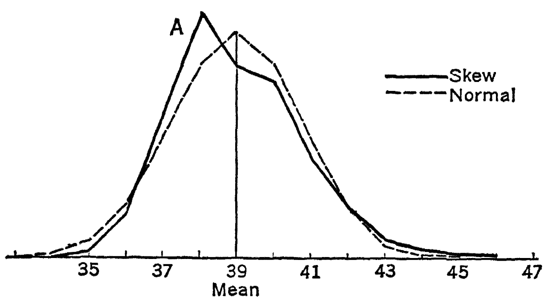


FIG. 5.1.—A distribution with positive skewness versus a normal distribution.

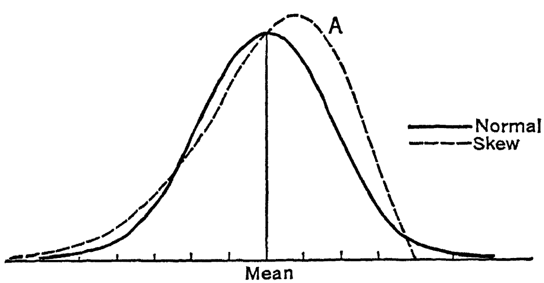


FIG. 5.2.—A frequency curve with negative skewness versus a normal frequency curve.

Figure 5.2 enables one to compare a normal frequency curve with a frequency curve possessing negative skewness. The skew curve has its maximum to the right of the mean and the tail to the left. Other situations may arise.

Skewness is often interpreted as the amount of deviation from the normal curve, since the skewness of the normal curve is zero. In reality, skewness measures the amount of asymmetry or the

amount of deviation from symmetry. Distributions of heights of people at various ages possess very little skewness; distributions of weights for various ages possess much skewness. Tails of weight distributions contain measurements for heavy people. Often distributions pertaining to number of seeds possess positive skewness because of the many more pods with large numbers of seeds than with small numbers. Frequency distributions with long tails to the left or to the right of the mean differ markedly from normal distributions.

It is not actually necessary to find deviations from the mean, as was done in Table 5.1, in calculating $\alpha_{3 \cdot v}$, for it can be found from the variates themselves. The average of the cubes of the deviations from the mean, or the third moment about the mean of the distribution, is

$$\mu_{3:v} = \frac{\Sigma(v_i - M)^3}{n} = \frac{\Sigma(v_i^3 - 3M \cdot v_i^2 + 3M^2 \cdot v_i - M^3)}{n}$$

$$= \frac{\Sigma v^3}{n} - \frac{3M \cdot \Sigma v^2}{n} + \frac{3M^2 \Sigma v}{n} - \frac{nM^3}{n},$$

$$M_{3:v} = \mu'_{3:v} - 3M \cdot \mu'_{2:v} + 2M^3.$$

Hence skewness is

$$\alpha_{3:v} = \frac{\mu'_{3:v} - 3M \cdot \mu'_{2 \cdot v} + 2M^3}{\sigma_v^3}, \tag{5.2}$$

which enables one to find $\alpha_{3:v}$ without finding deviations from the mean.

Since skewness is defined in terms of deviations of the variates from the mean it can be secured from the distribution obtained by subtracting a provisional mean from each variate; see Theorem 3.2. In other words,

$$\alpha_{3 \cdot v} = \alpha_{3:d} = \frac{\mu_{3:d}}{\sigma_d^3}, \tag{5.3}$$

which is easier to calculate since variates in the d-distribution are smaller than those in the original set. The next table contains a set of grouped variates for which skewness has been found, together with all necessary computations.

WEIGHTS OF FRESHMEN AT THE UNIVERSITY OF MICHIGAN

WEIGHTS	$f(v)$	d	$df(d)$	$d^2f(d)$	$d^3f(d)$
89.5- 99 5	1	−4	− 4	16	− 64
99.5–109.5	13	−3	− 39	117	−351
109 5–119.5	64	−2	−128	256	−512
119.5–129.5	128	−1	−128	128	−128
129 5–139.5	175	0	000	000	000
139.5–149.5	155	+1	155	155	+155
149.5–159.5	97	+2	194	388	776
159.5–169 5	37	3	111	333	999
169 5–179.5	27	4	108	432	1,728
179.5–189.5	9	5	45	225	1,125
189.5–199.5	7	6	42	252	1,512
199.5–209.5	2	7	14	98	686
	715		+370	2,400	5,926

$$M_v = 134.5 + (0.51748)10 = 139.6748 \text{ lb.},$$

$$\sigma_v = (\sqrt{3.35664 - 0.267786 - 0.083333})10 = 17.3364 \text{ lb.},$$

$$\mu_{3:d} = \mu'_{3:d} - 3M\cdot\mu'_{2:d} + 2M^3$$

$$= 8.28811 - 3(0.51748)(3.35664) + 2(.51748)^3$$

$$= 3.35417,$$

$$\alpha_{3:v} = \frac{\mu_{3\ d}\cdot 10^3}{(\sigma_{\text{adj.}})^3} = \frac{3.35417}{5.21047} = 0.6437.$$

Class marks of the distribution in the above table are exhibited graphically in Fig. 5.3. The part of the range to the right of the

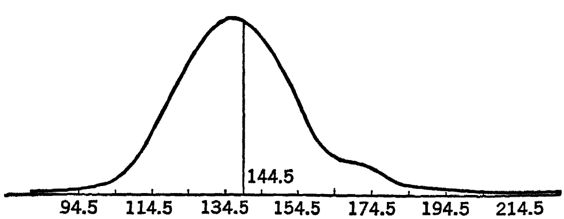


FIG. 5.3.—Frequency curve pertaining to weights of freshmen.

mean is larger than the part to the left. Notice that the maximum is to the left of the mean. About 69 per cent of this distribution lies within 1σ of the mean.

Grade curves are not normal, as was assumed in Chapter 4, for on examining a large number of grades one will soon discover no little amount of skewness. Percentages of grades at a large university for the year 1933–34 are given in Fig. 5.4. The frequency curve for this distribution is not normal; it possesses a rather large skewness. There are more A and B grades than D and E grades. Of course university students form a special population. One is not justified in using the normal distribution for university grades.

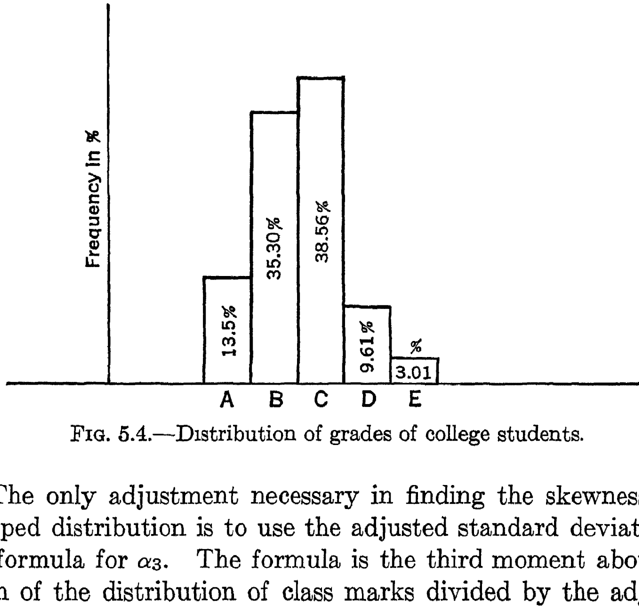


FIG. 5.4.—Distribution of grades of college students.

The only adjustment necessary in finding the skewness of a grouped distribution is to use the adjusted standard deviation in the formula for α_3. The formula is the third moment about the mean of the distribution of class marks divided by the adjusted standard deviation, or

$$\alpha_{3\ \text{adj.}} = \frac{\mu_{3\,\cdot\,\text{class marks}}}{(\sigma_{\text{adj.}})^3} = \frac{\mu_{3\,:\,d}\cdot w^3}{(\sigma_{\text{adj.}})^3}. \tag{5.4}$$

PROBLEMS

1. Using the percentages for grades given in Fig. 5.4, find the number of students in each grade group for 7,486 university students.

2. If $\alpha_{3\,:\,v} = 0.67$ and $\mu_{3\,:\,v} = 5.36$ cu. cm., find σ_v.

3. If $\alpha_{3\,:\,v} = 0.82$ and $\sigma_v = 14.21$, find $\Sigma \bar{v}^3/N$.

4. If $\alpha_{3\,:\,v} = 0.6$, $\sigma_v = 5$, $\Sigma \bar{v}^3 = 37{,}500$, find N.

5. Find the skewness of the distribution of rays in tail fins of flounders. See "Biometrika und Variationsstatistik," by P. Riebesell, page 760.

No. of Rays	No. of Flounders	No. of Rays	No. of Flounders
47	5	55	111
48	2	56	74
49	13	57	37
50	23	58	16
51	58	59	4
52	96	60	2
53	134	61	1
54	127		
			703

6. Find the skewness of the distribution of ages of women at marriage. Taken from *Detroit News* during 1936.

Ages at Marriage	No. of Women	Ages at Marriage	No. of Women
15–19	36	45–49	100
20–24	1,078	50–54	52
25–29	856	55–59	16
30–34	442	60–64	16
35–39	244	65–69	6
40–44	150	70–74	4
			3,000

Plot the distribution and discuss results.

7. Derive a formula for finding α_3 for combining two sets of data in terms of α_3's, σ's, M's, and N's.*

8. Use the formula derived in problem 7 for obtaining the skewness for the combination of the sets A and B if the following is known.

Set A: M_x = 139.55 lb., σ_x = 17.75 lb., $\alpha_{3:x} = 0.78$, $N_1 = 718$ = the number of items.
Set B: M_y = 138.65 lb., σ_y = 18.03 lb., $\alpha_{3:y} = 0.94$, $N_2 = 1{,}000$ = the number of items.

KURTOSIS

Some frequency curves are symmetrical and yet are not normal. Some possess more flatness at the mode than the normal; others are more peaked. Consider Figs. 5.5 and 5.6. Frequency curve A

* W. D. Baten, "A Formula for Finding the Skewness of the Combination of Two or More Examples," *Jour. Am. Stat. Assoc.*, 1935, pages 95–98.

in Fig. 5.5 deviates markedly from B, a normal curve, because of its flatness at the maximum. Curve C in Fig. 5.6 differs also from the normal B, because of its peak at the mode. Some symmetrical distributions have more items in the neighborhood of the mean than the normal, hence a tall peak at the mode.

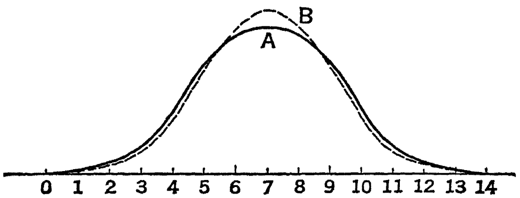


FIG. 5.5.—Normal distribution B and distribution with negative kurtosis A.

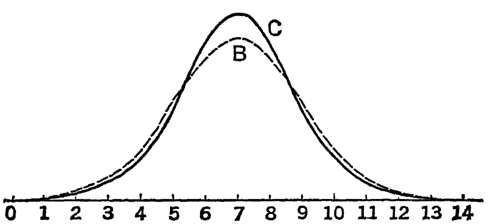


FIG. 5.6.—Normal distribution B and distribution with positive kurtosis C.

A measure of the amount of flatness or lack of flatness of a distribution is called kurtosis and is defined in terms of the average of the fourth powers of the deviations of the variates from the mean. This measure is

$$K = \alpha_{4:v} - 3 = \frac{\mu_{4 \cdot v}}{\sigma_v{}^4} - 3. \tag{5.5}$$

Kurtosis for normal curves is zero. Symmetrical frequency curves which are flatter at the mode than the normal curve have negative kurtosis; those which have a taller peak at the mode than the normal have positive kurtosis. Tables 5.2 and 5 3 contain respectively frequency distributions with negative and positive kurtosis, together with necessary computations for finding K. These distributions are shown in Figs. 5.5 and 5.6.

TABLE 5.2

v	$f(v)$	$\bar{v}$	$\bar{v}^2 f(v)$	$\bar{v}^4 \cdot f(v)$
0	1	−7	49	2,401
1	2	−6	72	2,592
2	11	−5	275	6,875
3	30	−4	480	7,680
4	73	−3	657	5,913
5	132	−2	528	2,112
6	166	−1	166	166
7	170	0	000	000
8	166	+1	166	166
9	132	2	528	2,112
10	73	3	657	5,913
11	30	4	480	7,680
12	11	5	275	6,875
13	2	6	72	2,592
14	1	7	49	2,401
	1,000		4,454	55,478

$M_v = 7;\ \sigma_v = \sqrt{4.454} = 2.11045,$

$\mu_{4:v} = 55.478,$

$\alpha_{4:v} = \dfrac{\mu_{4\cdot v}}{\sigma_v{}^4} = 2.79654,$

$K = \alpha_{4:v} - 3 = -0.20346.$

TABLE 5.3

v	$f(v)$	$\bar{v}$	$\bar{v}^2 \cdot f(v)$	$\bar{v}^4 \cdot f(v)$
0	1	−7	49	2,401
1	2	−6	72	2,592
2	8	−5	200	5,000
3	22	−4	352	5,632
4	54	−3	486	4,374
5	111	−2	444	1,776
6	190	−1	190	190
7	224	0	000	000
8	190	+1	190	190
9	111	2	444	1,776
10	54	3	486	4,374
11	22	4	352	5,632
12	8	5	200	5,000
13	2	6	72	2,592
14	1	7	49	2,401
	1,000		3,586	43,930

$M_v = 7;\ \sigma_v = \sqrt{3.586} = 1.89367,$

$\mu_{4:v} = 43.93,$

$\alpha_{4:v} = 3.416,$

$K = \alpha_{4\cdot v} - 3 = +0.416.$

It is not necessary to find the deviations of the variates from the mean in calculating kurtosis, for the fourth moment about the mean of a frequency distribution can be written as follows:

$$\mu_{4:v} = \frac{\Sigma(v - M)^4 f(v)}{\Sigma f(v)} = \frac{\Sigma v^4 f(v)}{\Sigma f(v)} - \frac{4M \cdot \Sigma v^3 f(v)}{\Sigma f(v)} + \frac{6M^2 \cdot \Sigma v^2 f(v)}{\Sigma f(v)} - 3M^4,$$

or

$$\mu_{4:v} = \mu'_{4:v} - 4M \cdot \mu'_{3:v} + 6M^2 \cdot \mu'_{2:v} - 3M^4. \tag{5.6}$$

From $\mu_{4:v}$ and σ_v, K can be obtained.

When data are grouped into classes it is necessary to use the following adjustments in calculating kurtosis.

For discrete variates*

$$(5.7)\ K = \frac{\left[\mu_{4:d} - (\tfrac{1}{2})(1-1/k^2)\mu_{2\ d} + \dfrac{(1-1/k^2)(7-3/k^2)}{240}\right]k^4}{(\sigma_{\text{adj.}})^4} - 3,$$

where k is the number of variates that can be put in a class, and d is the deviation from some provisional mean. For continuous variates

$$(5.8)\qquad K = \frac{[\mu_{4\ d} - \tfrac{1}{2}\mu_{2\ d} + \tfrac{7}{240}]w^4}{(\sigma_{\text{adj}})^4} - 3.$$

The following example will illustrate computations necessary for obtaining skewness and kurtosis for grouped data.

HIP MEASUREMENTS OF FRESHMEN AT THE UNIVERSITY OF MICHIGAN. MEASUREMENTS MADE TO THE NEAREST 0.1 INCH

Hip Measurement	Frequency	d	$df(d)$	$d^2f(d)$	$d^3f(d)$	$d^4f(d)$
25.95–27.95	2	−4				
27.95–29.95	3	−3				
29.95–31.95	7	−2				
31.95–33.95	92	−1				
33.95–35.95	222	0				
35.95–37.95	171	+1				
37.95–39.75	65	2				
39.95–41.96	13	3				
41.95–43.95	3	4				
43.95–45.95	2	5				
	580		+239	825	+1,127	5,241

$$M_v = h + M_d \cdot w = 34.95 + (0.412068)2 = 35.7741 \text{ in.},$$

$$\sigma_{\text{adj.}} = (\sqrt{1.42241 - 0.16980 - 0.08333})2 = 2.16266 \text{ in.},$$

$$\alpha_{3:v} = \frac{\mu_{3:d}\,2^3}{(\sigma_{\text{adj.}})^3} = \frac{(1.94310 - 1.75839 + 0.13994)8}{(2.16266)^3} = +\,0.25677,$$

$$K = \frac{(\mu_{4.d} - \tfrac{1}{2}\mu_{2.d} + \tfrac{7}{240})16}{(\sigma_{\text{adj}})^4} - 3 =$$

$$\frac{[(9.03621 - 3.20275 + 1.44915 - 0.086496) - 0.5(1.25261) + 0.029167]16}{(2.16266)^4} - 3$$

$$= +\,1.85316.$$

* H. C. Carver, "Editorial," *Annals of Mathematical Statistics*, Vol. I, No. 1, 1930, page 111, and J. R. Abernethy, "On the Elimination of Systematic Errors Due to Grouping," *ibid.*, Vol. IV, 1933, pages 263–277.

PROBLEMS

1. Find the skewness and kurtosis of the following distribution of weights of female babies at birth. Measurements made to nearest gram.

Weight at Birth	No. of Babies	Weight at Birth	No. of Babies
1,095–1,324	3	3,395–3,624	82
1,325–1,554	5	3,625–3,854	44
1,555–1,784	5	3,855–4,084	15
1,785–2,014	6	4,085–4,314	13
2,015–2,244	13	4,315–4,544	5
2,245–2,474	20	4,545–4,774	2
2,475–2,704	43	4,775–5,004	1
2,705–2,934	60		500
2,935–3,164	91		
3,165–3,394	92		

2. Find the kurtosis of the following distribution of stamens of flowers of May apples (*Podophyllum peltatum*) near Ann Arbor, Michigan.

No. of Stamens	No. of Flowers	No. of Stamens	No. of Flowers
9	3	15	90
10	30	16	41
11	95	17	8
12	369	18	3
13	246		1,000
14	115		

3. If $K = -0.12$, find $\alpha_{4:v}$.

4. Find K for the distribution on page 97.

CHAPTER 6

PERMUTATIONS, COMBINATIONS, AND PROBABILITY

INTRODUCTION

THEOREM 6.1. If a task can be done in m ways, and if after it has been done a second task can be done in n ways, then both tasks can be done, in this order, in mn ways.

PROOF: After the first task has been done the second task can be done in any one of the n ways. Since there are m ways of doing the first, there are mn ways of doing both tasks.

The meaning of this theorem will be illustrated by several examples.

EXAMPLE. 1. If there are 3 roads leading up the east side of a mountain and 5 down the west side, a person can cross the mountain on $3 \times 5 = 15$ different routes.

EXAMPLE 2. A coin and a 6-sided die can fall in 12 different ways, for after the coin has fallen the die might fall in 6 different ways.

Theorem 6.1 is very important, for it is used in much of the theory and in many of the problems to follow. The fundamentals of permutations, combinations, and probability are based upon this theorem.

EXAMPLE 3. A combination lock has 3 concentric rings of numbers. The lock is opened when 3 particular numbers, 1 on each ring, are in line; how many different combinations can be formed if the first ring contains the numbers 1, 2, 3, and 4, the second ring contains the numbers 3, 4, 5, 6, and 7 and the third ring contains the numbers 1 and 2?

According to the above theorem there will be $4 \times 5 \times 2 = 40$ different combinations.

PROBLEMS

1. In a classroom where boys and girls sit on opposite sides of the room there are 3 vacant seats on the girls' side and 5 on the boys' side. A boy and a girl enter the class late. In how many ways can they take seats?

2. In how many different ways can 3 ordinary dice fall?

3. In how many ways can an ordinary die and a 4-sided die fall?

4. In how many ways can a student post two letters if there are 5 mail boxes?

5. John Tol enters 4 track events and Ted Tol his brother enters 5 other events. How many first places may the Tol boys win? How many first and second places may they win?

6. There are 7 fraternities on a university campus. In how many ways may John and Ted choose fraternities?

7. There are 4 roads from A to B, 3 roads from B to C, and 5 from C to D. In how many different ways can a person go from A to D by way of B and C?

8. A student has 3 coats, 5 pairs of trousers, 3 hats, and 2 pairs of shoes. In how many ways can he dress?

9. A key has 5 teeth of 5 different lengths. How many different keys of 5 teeth can be made?

PERMUTATIONS

The different arrangements that can be formed from the letters a, b, and c taken all at a time are

$$abc,\ acb,\ bac,\ bca,\ cab, \text{ and } cba.$$

The arrangement acb is different from the arrangement bca. Order of the letters is important, for different orders give different arrangements.

Each of the arrangements that can be formed by taking some or all of a group of objects is called a permutation. The arrangement cab is also a permutation.

The permutations that can be formed from the letters a, b, and c taken 2 at a time are

$$ab,\ ba,\ ac,\ ca,\ bc,\ cb.$$

According to Theorem 6.1, the number of permutations that can be formed from 3 things taken 2 at a time is 3×2, since the first place can be filled in 3 ways and after it is filled the second place can be filled in 2 ways. Hence the number of permutations that can be formed from the letters a, b, and c taken 2 at a time is

$$_3P_2 = 3 \times 2 = 6 \text{ permutations.}$$

THEOREM 6.2. The number of permutations that can be formed from n different things taken r at a time is

$$_nP_r = n(n-1)(n-2)(n-3)\ldots(n-r+1). \tag{6.1}$$

PROOF: The first thing can be taken from any one of the n things; after the first has been chosen the second can be taken from any of the $n - 1$ remaining things; after the second has been chosen the third can be taken from any of the $n - 2$ remaining things, . . . , after the $(r - 1)$th has been chosen the rth thing may be chosen from any of the $n - r + 1$ remaining. Hence Theorem (6.2) follows from Theorem (6.1).

EXAMPLE 4. How many different 4-letter words can be formed from the letters in the word *thing*, where the same letter cannot be used twice in the same word? This can be obtained from (6.1); the number is the number of permutations that can be formed from 5 letters taken 4 at a time, and is

$${}_5P_4 = 5 \times 4 \times 3 \times 2 = 120 \text{ different ways.}$$

PROBLEMS

1. How many 3-digit numbers can be formed from the digits 1, 2, 3, 4, 5, 6, 7, 8, 9 if the same digit is not used twice in the number?

2. How many 3-digit numbers can be formed from the digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, if the same digit is not used twice in the number?

3. In how many ways can 6 people take seats in a coach with 9 vacant seats?

4. In how many ways may 10 different books be placed on a shelf?

5. How many different signals can be made with 5 different colored flags if there are 4 places for the flags and no color is to be used twice in the same signal?

6. How many different baseball teams can be formed from 10 men who can play in any position?

7. How many different baseball teams can be formed from 13 men, 2 of whom can only pitch, 3 of whom can only catch, 5 of whom can play infield, and 3 of whom can play outfield?

8. In how many ways can a man string 10 different colored beads?

9. A person is allowed to shoot 5 ducks, 3 rabbits, 2 pheasants, and 1 deer. How many ways can a person bring home 4 different kinds of game, where the limit was reached for each kind?

COMBINATIONS

In permutations, order was taken into consideration. The arrangement or permutation *ab* was different from the permutation *ba*. When order is not considered, the number of ways of selecting r things from n different things is called the number of combinations of n things taken r at a time. The selection or combination

ab is the same as *ba*. Permutations of the letters *a*, *b*, *c* taken 2 at a time are

$$ab,\ ba,\ ac,\ ca,\ bc,\ cb;$$

The combinations of these letters taken 2 at a time are

$$ab,\ ac,\ bc.$$

Since order does not enter into the number of combinations, the number of combinations can always be found from the number of permutations by dividing by the number of ways the letters can be interchanged in each arrangement. In the above example the number of combinations of the letters taken 2 at a time is equal to the number of permutations divided by 2!, or

$${}_3C_2 = \frac{{}_3P_2}{2!} = \frac{3 \times 2}{1 \times 2} = 3.$$

Theorem (6.3). The number of combinations of n different things taken r at a time is equal to the number of permutations of n things taken r at a time divided by factorial r, or

$$(6.2) \qquad {}_nC_r = \frac{{}_nP_r}{r!}.$$

Proof: Since order is not considered in finding the number of combinations the number of permutations divided by $r!$ will be equal to the number of combinations since there are $r!$ permutations of r things taken r at a time. In other words, for each way of selecting r things from n things there are $r!$ ways of arranging these r things. Hence the number of permutations is always $r!$ times as large as the number of combinations.

Example 5. How many different handshakes can 8 people make with each other? This is a combination problem, for when A shakes hands with B, B also shakes hands with A, and this is only 1 handshake. The answer to the question is the number of ways of taking 2 things from 8 different things, or the number of combinations that can be formed from 8 things taken 2 at a time; according to Theorem (6.3) it is equal to

$${}_8C_2 = \frac{{}_8P_2}{2!} = \frac{8 \times 7}{1 \times 2} = 28 \text{ handshakes.}$$

Example 6. How many different committees of 4 can be chosen from 7 people?

Here, again, order does not enter into the problem, for a committee of 4 is only 1 committee, regardless of how they sit in the committee room. The number of committees is the number of combinations that can be formed from 7 things taken 4 at a time, or

$$_7C_4 = \frac{7\cdot6\cdot5\cdot4}{1\cdot2\cdot3\cdot4} = 35 \text{ committees.}$$

Some of the committees will contain 3 men which were on other committees, but no 2 of the committees will contain the same 4 men.

EXAMPLE 7. How many triangles can be formed from 9 points, no three of which are in a straight line? This is a combination problem, for the triangle ABC is the same as the triangles ACB, CAB, BCA, etc. The answer is the number of combinations of 9 things taken 3 at a time, and is

$$_9C_3 = \frac{9\cdot8\cdot7}{1\cdot2\cdot3} = 84 \text{ different triangles.}$$

Sometimes a problem will involve permutations and combinations, as the following example illustrates.

EXAMPLE 8. How many words of 4 letters can be formed from 8 different letters so that the word does not have 2 letters alike?

There are $_8C_4 = \dfrac{8\cdot7\cdot6\cdot5}{1\cdot2\cdot3\cdot4} = 70$ ways of choosing groups of 4 letters from the 8 letters. After the 4 letters are chosen there are $4! = 24$ different words that can be formed from each group of 4 letters; hence the number of different words that can be formed from 70 groups of 4 letters is $70 \times 24 = 1{,}680$ different words, or $_8C_4\ 4! = 1{,}680$ different words.

This example might have been solved by finding the number of permutations of 8 things taken 4 at a time.

EXAMPLE 8a. There are 6 different consonants and 5 different vowels. How many words of 4 letters can be formed if the word is to have 2 consonants and 2 vowels and the same letter cannot be used twice in the same word?

There are $_6C_2$ ways of getting the 2 consonants and $_5C_2$ ways of getting the 2 vowels, hence there are $_6C_2\cdot{_5C_2}$ ways of getting the consonants and vowels for the words. After the 2 consonants and the 2 vowels have been chosen one can form 4! different words with them. Hence the number of 4-letter words is equal to

$$_6C_2\cdot{_5C_2}\cdot4! = \frac{6\cdot5}{1\cdot2}\cdot\frac{5\cdot4}{1\cdot2}\cdot1\cdot2\cdot3\cdot4 = 3{,}600 \text{ different words.}$$

Experience will enable one to know whether problems can be solved by formulas for permutations, combinations, or both.

PROBLEMS

1. How many straight lines can be formed from 11 points, no three of which are on a straight line?

2. How many ways can 3 white balls be drawn at random from a bag containing 5 white balls and 4 red balls if 3 balls are drawn at random?

3. How many committees of 7 students can be formed from 8 sophomores and 5 freshmen if the committee is to consist of 4 sophomores and 3 freshmen?

4. How many basketball teams can be formed from 6 men if each can play in any position?

5. How many different examinations of 6 questions can be formed from 10 different questions? How many of these will contain the first 3 questions?

6. A bag contains 4 white balls, 5 red balls, and 7 blue balls. A person draws at random 5 balls. How many ways can he draw?

(*a*) 2 white and 3 red balls?
(*b*) 5 blue balls?
(*c*) 1 white, 2 red, and 2 blue balls?
(*d*) 5 red balls?
(*e*) 2 white, 1 red, and 2 blue balls?

7. How many different services can be held with 3 preachers, 6 singers, 7 deacons and 2 organists if a service requires 1 preacher, 4 singers, 3 deacons, and 1 organist?

8. How many parallelograms can be formed from 7 vertical parallel lines and 5 horizontal parallel lines?

9. How many ways may 13 cards be drawn from a deck of 52 cards?

10. A railway line has 12 stations. How many different tickets must be made so that each station can sell a ticket to any other station?

11. How many different sums of money can be formed with a penny, a nickel, a dime, a quarter, a half dollar, and a dollar?

12. How many of the sums in problem 11 do not contain the dime and half dollar?

PROBABILITY

If an event can happen in m different ways and fail in n different ways and all of the ways are equally likely, then the probability of the event happening on any trial is

$$p = \frac{m}{m + n}; \tag{6.3}$$

the probability of the event not happening is

$$q = \frac{n}{m+n}. \tag{6.4}$$

According to this definition the probability of an event happening on any trial is the number of favorable ways to its occurrence divided by the total number of ways in which it can happen and fail.

EXAMPLE 9. A man draws at random a ball from a bag containing 6 red balls and 4 purple balls, all of which are of the same size. According to the above definition the probability of drawing a purple ball is

$$p = \frac{4}{4+6} = \frac{4}{10},$$

since the purple ball can be drawn in 4 ways out of 10 possible ways of drawing the ball.

PROBLEMS

1. What is the chance of getting the sum of the spots to turn up on 2 dice to equal 5? To equal 11? 10? 7?

2. Out of a bag containing 5 red balls and 4 black balls there are drawn at random 3 balls. What is the probability that:

(*a*) All will be red?
(*b*) 2 will be red and 1 black?
(*c*) 2 will be green?
(*d*) 2 will be black?
(*e*) 1 will be red and 2 black?

3. What is the probability of getting a 4-spot on a 4-sided die and a 5-spot on a 6-sided die if both are thrown? What is the probability of getting the sum of the spots to be equal to 7? 5? 10?

4. The probability of house A burning is 0.08, and the probability of house B burning is 0.04. What is the probability of:

(*a*) both houses burning?
(*b*) A not burning and B burning?
(*c*) A and B not burning?

5. At a party there are 24 people and 5 prizes. Out of a bag containing slips of paper with the names of the people present 1 slip is drawn at random and is put back in the bag before the next drawing. What is the probability that:

(*a*) Mr. Jones will get a prize?
(*b*) Mr. Jones will receive all the prizes?
(*c*) Mr. and Mrs Jones will receive no prize?
(*d*) Mr. and Mrs. Jones will each get a prize?
(*e*) Mr. and Mrs. Jones will receive all the prizes?

6. The following histogram of a distribution containing 1,000 items is given.

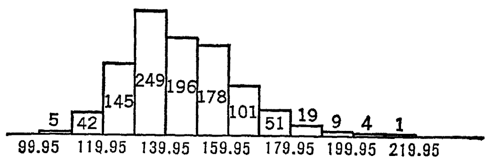


FIG. 6.0.—Histogram of a distribution of weights of 1,000 men.

If an item is drawn at random, what is the probability that it will lie between 139.95 and 159.95? between 99.95 and 119.95? between 142 and 154? between 163.6 and 170.3?

7. On a slot machine there are 3 disks with the 10 digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, on each. When a coin is inserted the 3 disks revolve several times and come to rest. What is the probability of getting 3 particular numbers, one on each disk, to appear?

PROBABILITY IN REPEATED TRIALS

If the probability of success in 1 trial is $\frac{3}{4}$ and the probability of failure is $\frac{1}{4}$, then the probability of 2 successes in 3 trials may be analyzed as follows:

1. A success on the first and second trials and a failure on the third; the probability of such is $(\frac{3}{4})(\frac{3}{4})(\frac{1}{4}) = \frac{9}{64}$.

2. Successes on the first and third trials and failure on the second; the probability of such is $(\frac{3}{4})(\frac{1}{4})(\frac{3}{4}) = \frac{9}{64}$.

3. Failure on the first trial and success on the last two trials; the probability of such is $(\frac{1}{4})(\frac{3}{4})(\frac{3}{4}) = \frac{9}{64}$.

Hence the probability of 2 successes in 3 trials is the sum of these probabilities, or $\frac{9}{64} + \frac{9}{64} + \frac{9}{64} = \frac{27}{64}$.

This might have been found by taking the third term in the expansion of the binomial $(\frac{1}{4} + \frac{3}{4})^3$, for

$$(\tfrac{1}{4} + \tfrac{3}{4})^3 = (\tfrac{1}{4} + \tfrac{3}{4})(\tfrac{1}{4} + \tfrac{3}{4})(\tfrac{1}{4} + \tfrac{3}{4}),$$

$$(6.5) \qquad (\tfrac{1}{4} + \tfrac{3}{4})^3 = (\tfrac{1}{4})^3 + 3(\tfrac{1}{4})^2(\tfrac{3}{4}) + 3(\tfrac{1}{4})(\tfrac{3}{4})^2 + (\tfrac{3}{4})^3.$$

The right-hand side of (6.5) was obtained by multiplying the 3 factors in the above line together. The third term in the right-hand member of (6.5) was obtained by adding the following 3 products of 3 factors: $\frac{3}{4}$ in the first factor by $\frac{3}{4}$ in the second factor by $\frac{1}{4}$ in the third factor; $\frac{3}{4}$ in the first factor, by $\frac{1}{4}$ in the second factor by $\frac{3}{4}$ in the third factor; $\frac{1}{4}$ in the first factor, by $\frac{3}{4}$ in the second factor by $\frac{3}{4}$ in the third factor.

The last term in (6.5) is the product of $(\frac{3}{4})(\frac{3}{4})(\frac{3}{4})$ and is the probability that there will be a success on the first, second, and third trials. The second term in (6.5) is the probability that there will be 1 success and 2 failures in the 3 trials, while the first term is the product $(\frac{1}{4})(\frac{1}{4})(\frac{1}{4})$, which is the probability of a failure on each of the 3 trials.

Suppose that a die is thrown 4 times and a success is considered getting an ace to turn up. The terms in the expansion of $(\frac{1}{6} + \frac{5}{6})^4$ give the probabilities of the various numbers of successes. They may be exhibited as follows:

$$(6.6)\qquad (\tfrac{5}{6} + \tfrac{1}{6})^4 = \overbrace{(\tfrac{5}{6})^4 + {}_4C_1(\tfrac{5}{6})^3(\tfrac{1}{6}) + {}_4C_2(\tfrac{5}{6})^2(\tfrac{1}{6})^2 + {}_4C_3(\tfrac{5}{6})^1(\tfrac{1}{6})^3 + {}_4C_4(\tfrac{1}{6})^4}^{\text{Probability of}}$$

Successes and	0	1	2	3	4
failures	4	3	2	1	0

The first term in the right member of (6.6) is the probability of no successes and 4 failures in the 4 trials, the second term is the probability of 1 success and 3 failures, the third term is the probability of 2 successes and 2 failures, the fourth term is the probability of 3 successes and 1 failure, and the last term is the probability of 4 successes in the 4 trials. The reason the terms of the binomial expansion can be used for the probabilities is that the coefficients of the terms give the number of ways the various successes and failures may occur in the 4 trials. For example, consider the probability of getting 1 success in 4 trials; there may be 1 success on the first trial and failure on the other trials, or 1 success on the second trial and failure on the others, or a success on the third trial and failure on the others, or a success on the last trial and failure on the others; this gives ${}_4C_1$ different ways for 1 success in the 4 trials. The sum of probabilities for these 4 will give the probability of getting 1 success in the 4 trials. This sum is

$$(\tfrac{1}{6})(\tfrac{5}{6})^3 + (\tfrac{1}{6})(\tfrac{5}{6})^3 + (\tfrac{1}{6})(\tfrac{5}{6})^3 + (\tfrac{1}{6})(\tfrac{5}{6})^3 = 4(\tfrac{1}{6})(\tfrac{5}{6})^3,$$

which is the second term in (6.6). Other terms of (6.6) may also be explained.

In general, if the probability of occurrence of an event in each trial is p and the probability of its non-occurrence is q and the trials are independent, then the probabilities of the various numbers of successes in n trials are given by the terms in the expansion of the binomial $(q + p)^n$, or

$$(6.7) \quad (q + p)^n = {}_nC_0 q^n + {}_nC_1 q^{n-1}p + {}_nC_2 q^{n-2}p^2 + {}_nC_3 q^{n-3}p^3 + \ldots + {}_nC_r\, q^{n-r}p^r + \ldots + {}_nC_{n-1} q \cdot p^{n-1} + {}_nC_n p^n,$$

where the first term is the probability of no successes, the second term is the probability of 1 success and $n - 1$ failures, the third term is the probability of 2 successes and $n - 2$ failures, etc. The $(r + 1)$th term is the probability of r successes and $n - r$ failures in n trials; for the r successes might happen in ${}_nC_r$ different ways, that is, there are ${}_nC_r$ trials among the n trials in which there may be r success and the rest failures.

Example 9a. An ordinary die is thrown 5 times, or 5 ordinary dice are thrown once. The probabilities for the various successes are given by terms in the expansion of $(\frac{5}{6} + \frac{1}{6})^5$, since the probability of getting an ace on each trial is $\frac{1}{6}$ and the probability of not getting an ace is $\frac{5}{6}$. This expansion is

$$(\tfrac{5}{6} + \tfrac{1}{6})^5 = (\tfrac{5}{6})^5 + 5(\tfrac{5}{6})^4(\tfrac{1}{6}) + 10(\tfrac{5}{6})^3(\tfrac{1}{6})^2 + 10(\tfrac{5}{6})^2(\tfrac{1}{6})^3 + 5(\tfrac{5}{6})\,(\tfrac{1}{6})^4 + (\tfrac{1}{6})^5.$$

The probability of getting 3 aces in the 5 trials is given by the fourth term in the expansion and is $10\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)^3 = \frac{250}{7{,}776}$.

The probability of getting exactly 1 ace in 5 throws is given by the second term in the expansion and is $5\left(\frac{5}{6}\right)^4\left(\frac{1}{6}\right) = \frac{3{,}125}{7{,}776}$.

The probability of getting at least 4 aces is given by adding the last 2 terms and is $5\left(\frac{5}{6}\right)\left(\frac{1}{6}\right)^4 + \left(\frac{1}{6}\right)^5 = \frac{25}{7{,}776} + \frac{1}{7{,}776} = \frac{26}{7{,}776}$.

The probability of getting at least 2 and no more than 3 aces is given by adding the third and fourth terms and is

$$10\left(\frac{5}{6}\right)^3\left(\frac{1}{6}\right)^2 + 10\left(\frac{5}{6}\right)^2\left(\frac{1}{6}\right)^3 = \frac{1{,}250}{7{,}776} + \frac{250}{7{,}776} = \frac{1{,}500}{7{,}776}.$$

PROBLEMS

1. A 4-sided die is thrown 6 times. Find the probability of getting:

(*a*) exactly 4 aces;
(*b*) at least 3 aces;
(*c*) exactly 6 aces;
(*d*) at most 2 aces;
(*e*) at least 3 and at most 5 aces.

2. The probability of an event happening in 1 trial is $\frac{2}{3}$. Find the probability, in 7 trials, of getting:

(*a*) exactly 4 successes;
(*b*) at least 5 and no more than 6 successes;
(*c*) less than 3 successes;
(*d*) 5 failures;
(*e*) all failures;
(*f*) not 5 successes.

3. A coin is tossed 10 times. Find the probability of getting:

(*a*) exactly 2 heads;
(*b*) exactly 6 tails;
(*c*) more than 7 tails.
(*d*) 9 heads and 1 tail.

4. The probability of any ship of a company being destroyed on a certain voyage is 0.02. The company owns 6 ships for this voyage. What is the probability of: (*a*) losing 1 ship? (*b*) losing at most 2 ships? (*c*) losing none?

5. The probability of its raining $\frac{1}{2}$ inches on any of the 3 days before Christmas is 0.08 for a certain town. A merchant takes out rain insurance for these 3 days. Find the probability that (*a*) it will not rain during the 3 days; (*b*) $\frac{1}{2}$ inch of rain will fall during 1 of the 3 days; (*c*) it will rain $\frac{1}{2}$ inch only on the last day; (*d*) it will rain $\frac{1}{2}$ inch every day.

6. If a man hits a traget 3 times out of 5, find the probability that he will, on 4 shots, hit the target; (*a*) exactly 3 times; (*b*) 1 time only; (*c*) no time; (*d*) at least twice.

7. The mean and standard deviation of a normal distribution of heights of boys are respectively 47.8 inches and 2.1 inches. A boy is picked at random from the distribution of 3,489 heights. (Original measurement made to the nearest 0.25 inch.) Find the probability that his height is (*a*) greater than 51.5 inches; (*b*) at least 49 inches; (*c*) between 45.25 inches and 50.75 inches and not including either; (*d*) less than 44 inches; (*e*) either between 44.75 inches and 46 inches inclusively or

between 52 inches and 54.5 inches, not to include 52 and 54.5; (*f*) exactly 50 inches—use ordinate table here.

8. The mean and standard deviation of a normal distribution of 2,484 measurements are respectively 20.6 centimeters and 1.2 centimeters. A measurement is picked at random. The probability of this measurement being between 21 centimeters and w centimeters inclusive is 0 2268. Find the number w if measurements were made to the nearest unit.

9. For a normal distribution find the probability that an item that is picked at random is within 1 σ of the mean. Between $+2$ σ's and $+3$ σ's of the mean.

10. For any distribution, what is the probability that an item picked at random is between the first and third quartiles? Between the third quartile and the extreme right end of the range?

CHAPTER 7

BERNOULLI DISTRIBUTION

FRACTION FREQUENCIES

Let the frequencies of a given distribution be multiplied by a constant H, thus forming a new distribution with total frequency H times the total frequency of the original distribution. The two distributions have variates identically equal, but the frequencies of the new distribution are H times as large as the corresponding frequencies of the original. How are the characteristics, that is, M, σ, α_3, and K, of the new and original distributions related?

Let the original set of variates be given by the first two columns in the table below, and the variates with new frequencies be given by the third and fourth columns. The characteristics of the new distribution will be found and compared with those of the original.

v	$f(v)$	v	$Hf(v)$	$vHf(v)$	$v^2Hf(v)$	$v^3Hf(v)$
v_1	$f(v_1)$	v_1	$Hf(v_1)$	$v_1Hf(v_1)$	$v_1{}^2Hf(v_1)$	$v_1{}^3Hf(v_1)$
v_2	$f(v_2)$	v_2	$Hf(v_2)$	$v_2Hf(v_2)$	$v_2{}^2Hf(v_2)$	$v_2{}^3Hf(v_2)$
v_3	$f(v_3)$	v_3	$Hf(v_3)$	$v_3Hf(v_3)$	$v_3{}^2Hf(v_3)$	$v_3{}^3Hf(v_2)$
.	.	.	.	.	.	.
.	.	.	.	.	.	.
v_n	$f(v_n)$	v_n	$Hf(v_n)$	$v_nHf(v_n)$	$v_n{}^2Hf(v_n)$	$v_n{}^3Hf(v_n)$

The mean of the new distribution is

$$M_{\text{new}} = \frac{v_1Hf(v_1) + v_2Hf(v_2) + \ldots + v_nHf(v_n)}{Hf(v_1) + Hf(v_2) \ldots + Hf(v_n)} = \frac{H\Sigma vf(v)}{H\Sigma f(v)} = M_v.$$

In other words, the mean of a new distribution formed by multiplying the frequencies of a given distribution by a constant is equal to the mean of the original distribution.

The second moment of the data with new frequencies is

$$\mu'_{2:\text{new}} = \frac{v_1^2Hf(v_1) + v_2^2Hf(v_2) + \ldots + v_n^2Hf(v_n)}{Hf(v_1) + Hf(v_2) + \ldots + Hf(v_n)} = \frac{H\Sigma v^2 f(v)}{H\Sigma f(v)}$$

$$= \mu'_{2\ \cdot v}, \text{ the second moment of the original distribution.}$$

The standard deviation of the distribution with frequencies H times as large as the given frequencies is

$$\sigma_{\text{new}} = \sqrt{\mu'_{2\ \text{new}} - (M_{\text{new}})^2} = \sqrt{\mu'_{2\cdot v} - M_v^2} = \sigma_v.$$

In a similar way it can be shown that α_3 and K are respectively equal for the two distributions.

The above has shown that if two distributions have identical variates, but the frequencies of the second are H times as large as the corresponding frequencies of the first, then the characteristics of the two are respectively equal.

If H is a fraction the frequencies of the new distribution might be fractions instead of integers; this would not affect M, σ, α_3, and K. If $H = 1/N$, where N is the total frequency of the given set, then the frequencies of the new distribution are in percentages of the total frequencies of the given set. In this case, frequencies are the probabilities of the various variates being drawn when one variate is drawn at random. Tables I and II contain distributions with frequencies expressed in percentages; each frequency is a probability. Consider the set of variates in the first two columns of the following table. Let a second distribution have the same variates and frequencies equal to $H = 1/100$ times as large as the frequencies in column 2. These are listed in column 3.

		$H = 0.01$
v	$f(v)$	$0.01 \cdot f(v)$
67	2	0.02
68	3	.03
69	6	.06
70	12	.12
71	21	.21
72	17	.17
73	13	.13
74	8	.08
75	5	.05
76	2	.02
77	1	.01
	100	1.00

Columns 1 and 3 form a new distribution with total frequency equal to unity and with M, σ, α_3, and K equal to those of the distribution in columns 1 and 2.

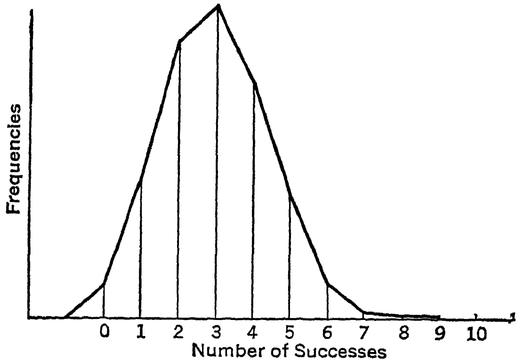


FIG. 6 1.—Probability or frequency distribution of the number of successes in nine trials when the probability of success in each trial is equal to $\frac{1}{3}$.

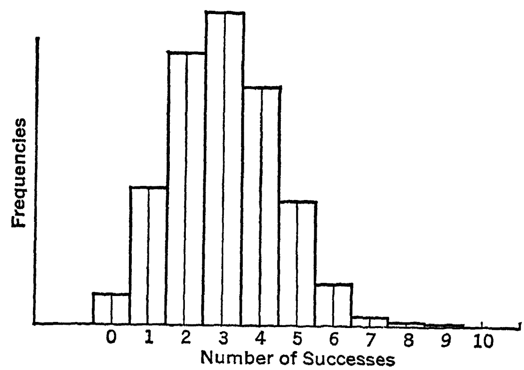


FIG. 6.2.—Histogram of the probability or frequency distribution of the number of successes in nine trials when the probability of success in each trial is $\frac{1}{3}$.

The above shows that frequency distributions may have fractions for frequencies.

Let the probability of an event's happening in each trial be $\frac{1}{3}$ and the probability of its not happening in each trial $\frac{2}{3}$. In 9 trials the event may happen no times, 1 time, 2 times, . . . , 8 times, or 9 times. The probabilities or frequencies of the various numbers of occurrences are given by the terms in the expansion of the

binomial $(\frac{2}{3} + \frac{1}{3})^9$; these are arranged in a frequency distribution in the following table, where v represents the various numbers of occurrences in the 9 trials.

v	Frequency $f(v)$
0	$(\frac{2}{3})^9 = 512/19{,}683$
1	$9(\frac{2}{3})^8(\frac{1}{3}) = 2{,}304/19{,}683$
2	$36(\frac{2}{3})^7(\frac{1}{3})^2 = 4{,}608/19{,}683$
3	$84(\frac{2}{3})^6(\frac{1}{3})^3 = 5{,}376/19{,}683$
4	$126(\frac{2}{3})^5(\frac{2}{3})^4 = 4{,}032/19{,}683$
5	$126(\frac{2}{3})^4(\frac{1}{3})^5 = 2{,}016/19{,}683$
6	$84(\frac{2}{3})^3(\frac{1}{3})^6 = 672/19{,}683$
7	$36(\frac{2}{3})^2(\frac{1}{3})^7 = 144/19{,}683$
8	$9(\frac{2}{3})\ (\frac{1}{3})^8 = 18/19{,}683$
9	$\ldots.\ (\frac{1}{3})^9 = 1/19{,}683$
	$\Sigma f(v) = 1$

The sum of the frequencies in the above table is unity, the mean is 3 occurrences, and the standard deviation is 1.414 occurrences. This distribution is graphically represented in Figs. 6.1 and 6.2.

MOMENTS OF BERNOULLI DISTRIBUTION

In general, if p is the probability of the occurrence of an event in each trial and q is the probability of the non-occurrence of the event in each trial and the trials are independent, then the frequencies of the possible numbers of occurrences in n trials are given by the terms of the expansion of the binomial $(q + p)^n$. Such a distribution is called a Bernoulli distribution. This distribution is given in Table 7.1 The characteristics of the general Bernoulli frequency distribution will now be derived.

The mean of the Bernoulli distribution in Table 7.1 is found by summing the third column; after factoring out of each term np, this sum is

$$M_v = np\left[q^{n-1} + (n-1)q^{n-2}p^1 + \frac{(n-1)(n-2)}{2!}q^{n-3}p^2 + \ldots + (n-1)qp^{n-2} + p^{n-1}\right]$$

$$= np(q + p)^{n-1} = np, \text{ since } (q + p)^{n-1} = 1.$$

This shows that the mean of a Bernoulli distribution is equal to the

TABLE 7.1

BERNOULLI DISTRIBUTION

v	$f(v)$	$v \cdot f(v)$	$v(v-1)f(v)$	$v(v-1)(v-2)f(v)$
0	q^n	0	0	0
1	$nq^{n-1}p$	$1 \cdot nq^{n-1}p$	0	0
2	$\frac{n(n-1)}{2!} q^{n-2}p^2$	$\frac{2\, n(n-1)}{2!} q^{n-2}p^2$	$\frac{2 \cdot 1n(n-1)}{2!} q^{n-2}p^2$	0
3	$\frac{n(n-1)(n-2)}{3!} q^{n-3}p^3$	$\frac{3n(n-1)(n-2)}{3!} q^{n-3}p^3$	$\frac{3 \cdot 2n(n-1)(n-2)}{3!} q^{n-3}p^3$	$\frac{3 \cdot 2 \cdot 1n(n-1)(n-2)}{3!} q^{n-3}p^3$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
$n-1$	nqp^{n-1}	$(n-1)nqp^{n-1}$	$(n-1)(n-2)nqp^{n-1}$	$(n-1)(n-2)(n-3)nqp^{n-1}$
n	p^n	np^n	$n(n-1)\, p^n$	$n(n-1)(n-2)\, p^n$
Total	$1 = \Sigma f(v)$	np	$n^2p^2 - np^2$	$n^3p^3 - 3n^2p^3 + 2np^3$

number of trials, n, multiplied by p, the probability of the occurrence of the event in each trial.

The second moment of this distribution is found from column 4. The sum of the quantities in this column is

$$\Sigma v(v-1)f(v) = \Sigma v^2 f(v) - \Sigma v f(v) = n(n-1)p^2 \Big[q^{n-2} + (n-2)q^{n-3}p^1 + \frac{(n-2)(n-3)}{2!} q^{n-4}p^2 + \ldots + (n-2)qp^{n-3} + p^{n-2} \Big]$$

$$= n(n-1)p^2(q+p)^{n-2} = n(n-1)p^2 = n^2p^2 - np^2.$$

$$\therefore\ \Sigma v^2 \cdot f(v) = n^2p^2 - np^2 + \Sigma v f(v).$$

Hence

$$\mu'_{2:v} = \Sigma v^2 f(v) = n^2p^2 - np^2 + \Sigma v f(v) = n^2p^2 - np^2 + np,$$

since $\Sigma v f(v) = np$ as found for the mean. Therefore

$$\mu_{2:v} = \mu'_{2.v} - M_v^2 = n^2p^2 - np^2 + np - n^2p^2 = np(1-p) = npq,$$

since $q + p = 1$. The standard deviation is

$$\sigma_v = \sqrt{npq},$$

which shows that the standard deviation is the square root of the product of n, the number of independent trials; p, the probability of the occurrence of the event in each trial; and q, the probability of failure in each trial.

The third moment can be obtained from column 5, for the sum of the numbers in this column is

$$\Sigma v(v-1)(v-2)f(v) = \Sigma v^3 f(v) - 3\Sigma v^2 f(v) + 2\Sigma v f(v)$$

$$= n(n-1)(n-2)p^3 [q^{n-3} + (n-3)q^{n-4}p + \frac{(n-3)(n-4)}{2!} q^{n-5}p^2 + \ldots + (n-3)qp^{n-4} + p^{n-3}]$$

$$= n(n-1)(n-2)p^3(q+p)^{n-3}$$

$$= n^3p^3 - 3n^2p^3 + 2np^3.$$

Hence

$$\begin{aligned}\mu'_{3\,.\,v} = \Sigma v^3 f(v) &= n^3p^3 - 3n^2p^3 + 2np^3 + 3\Sigma v^2 f(v) - 2\Sigma v f(v)\\ &= n^3p^3 - 3n^2p^3 + 2np^3 + 3n^2p^2 - 3np^2 + 3np - 2np\\ &= n^3p^3 - 3n^2p^3 + 2np^3 + 3n^2p^2 - 3np^2 + np.\end{aligned}$$

The third moment about the mean is

$$\begin{aligned}\mu_{3\,:\,v} &= \mu'_{3\,.\,v} - 3M_v \cdot \mu'_{2\,.\,v} + 2M_v^3\\ &= n^3p^3 - 3n^2p^3 + 2np^3 + 3n^2p^2 - 3np^2 + np\\ &\qquad - 3np(n^2p^2 - np^2 + np) + 2n^3p^3\\ &= np(2p^2 - 3p + 1) = np[2p(p-1) + (1-p)],\end{aligned}$$

or

$$\mu_{3\,:\,v} = np(-2pq + q) = npq(q - p).$$

The skewness is

$$\alpha_{3\,:\,v} = \frac{\mu_{3\,:\,v}}{\sigma^3} = \frac{q-p}{\sqrt{npq}}.$$

$$\alpha_{3\,:\,v} \gtreqless 0 \text{ according as } q \gtreqless p.$$

The mean, standard deviation, and skewness of a Bernoulli distribution are respectively

$$M_v = np, \quad \sigma = \sqrt{npq}, \quad \alpha_3 = \frac{q-p}{\sqrt{npq}}.$$

APPLICATION

EXAMPLE. If 720 ordinary dice are thrown on the table, what is the expected number of aces?

The probability of an ace for each die or trial is $\frac{1}{6} = p$, while $q = \frac{5}{6}$. Hence the expected number of aces or mean number of aces is $np = 720(\frac{1}{6}) = 120$. This means that on the average 120 aces will be found among the faces to turn up. This does not mean that exactly 120 aces will turn up every time 720 dice are thrown on the table, but that in a long series of throws with 720 dice the average number of aces will be 120.

When 720 dice are thrown the possible numbers of aces to turn up range from 0 to 720. The frequencies or probabilities of these various numbers of aces are given by the terms in the expansion of the binomial $(\frac{5}{6} + \frac{1}{6})^{720}$. The standard deviation of this Bernoulli distribution is $\sigma = \sqrt{720(\frac{5}{6})(\frac{1}{6})} = 10$ aces, and the skewness is $\dfrac{\frac{5}{6} - \frac{1}{6}}{10} = 0.067$, which is not far from the skewness of a normal distribution.

EXAMPLE. Let it be required to find the probability of getting more than 140 aces in 1 throw of 720 dice, also the probability of getting at least 112 aces and no more than 133 aces.

The probability for the first part is the sum of the 580 terms in the expansion of the binomial $(\frac{5}{6} + \frac{1}{6})^{720}$ beyond the term that gives the probability of exactly 140 aces, or is the sum of the terms

$$_{720}C_{141}(\tfrac{5}{6})^{579}(\tfrac{1}{6})^{141} + {}_{720}C_{142}(\tfrac{5}{6})^{578}(\tfrac{1}{6})^{142} + \ldots + (\tfrac{1}{6})^{720}.$$

It would be very difficult to find this sum by arithmetic or by means of logarithms. An approximation to it can be found by employing the normal curve in Table I, since the skewness of this Bernoulli distribution differs very little from zero. We want to find the area under the normal curve beyond the t that corresponds to 140.5 (we are dealing with discrete variates). The corresponding t is

$$t = \frac{140.5 - 120}{10} = +2.05 \text{ standard units.}$$

The area to the left of this value of t is found from Table I to be 0.97982; hence the area to the right of this t is 0.02018, which is the approximate probability of getting more than 140 aces on 1 throw of 720 dice. This means that if one threw the 720 dice down 100 times one could expect more than 140 aces to appear 2 times.

The t's that correspond to the second part of the above example are

$$t_1 = \frac{111.5 - 120}{10} = -0.85, \qquad t_2 = \frac{133.5 - 120}{10} = 1.35.$$

The area between these two t's under the normal curve is found from Table I to be 0.71383; hence the probability of getting at least 112 and no more than 133 aces is 0.71383, which means that 71 times out of 100 throws with 720 dice one can expect between 112 and 133 aces to turn up.

EXAMPLE. The probability of a man of age A dying within a year is 0.02. An insurance company has 10,000 men of this age insured with it. Find the probability of having to pay less than 180 death claims; also the probability of having to pay exactly 217 death claims.

This is solved by means of a Bernoulli distribution, for the event, a man dying within a year, has probability of 0.02 in each "trial" and 0.98 for failure. The mean, standard deviation, and skewness of the distribution made up of the possible numbers of men dying within the year are:

$$M_v = 200 \text{ men}, \quad \sigma_v = 14 \text{ men}, \quad \alpha_v = 0.07; \text{ call it zero.}$$

The t that corresponds to the lower limit of the rectangle in the histogram which contains the 180th item is

$$t = \frac{179.5 - 200}{14} = -1.46 \text{ standard units.}$$

The area under the normal curve to the left of this t is 0.07215, hence the probability of having to pay less than 180 death claims is 0.07215. This means that if the company had 100 groups of 10,000 men it would expect to have less than 180 death claims to pay in 7 of these groups, or if 100 companies each had 10,000 men of age A, 7 of these companies would on the average pay less than 180 death claims during the year.

The second part of the above example will be solved by means of the ordinate table for a normal curve. The t corresponds to the class mark of the group that contains 217 is

$$t = \frac{217 - 200}{14} = 1.21 \text{ standard units.}$$

The ordinate of the normal curve with unity frequency is found from Table II to be 0.19186; hence according to formula (4.5) the ordinate of this Bernoulli distribution for the v that corresponds to this t is

$$f(v) = \frac{1}{\sigma} f(t) = \frac{1}{14} (0.19186) = 0.01370,$$

which is the probability of having to pay exactly 217 claims during the year.

This may also be found by the area Table I for

$$t_1 = \frac{216.5 - 200}{14} = 1.18, \quad t_2 = \frac{217.5 - 200}{14} = 1.25.$$

The area between these t's under this normal curve is 0.01335, which is the probability of there being exactly 217 deaths during the year. The result is nearly the same as found by the ordinate table. For finding the probability for an exact number of events the ordinate table is preferred; however, there is very little difference between them in many cases.

PROBLEMS

1. If the probability of an event happening in each trial is $\frac{1}{4}$, find the probability of the event happening 3 times in 8 trials. If v represents the numbers of possible occurrences in 8 trials, plot the histogram and frequency polygon of this distribution.

2. A coin is thrown 10 times. Find the probability of getting:

(*a*) 5 heads to turn up; (*b*) 7 tails; (*c*) Plot the frequency polygon and histogram of the distribution.

3. Twelve hundred 4-sided dice are thrown on the table. Find the expected number of aces to turn down. Find the probability of getting:

(*a*) more than 324 aces;
(*b*) less than 281 aces;
(*c*) at least 281 aces and at most 324 aces;
(*d*) exactly 300 aces.
(*e*) express the answer for (*d*) as a term in a binomial expansion and try to find its value.

4. Forty-nine men of a certain age, out of 1,000 men of this age, died during the last year. An insurance company has 3,482 men of this age insured with it. Find the number of death claims the company expects to pay during the year. Find the probability of paying:

(*a*) more than 175 death claims;
(*b*) less than 166 claims;
(*c*) exactly 170 death claims.

5. If a new distribution is formed by multiplying each variate of a given distribution by a constant H, how are the characteristics of the new distribution related to those of the given distribution?

6. Find the kurtosis of the general Bernoulli distribution. What does this approach when $n \to \infty$?

CHAPTER 8

INDEX NUMBERS

RELATIVES

Consider prices of tobacco from 1924 to 1934 given in Table 8.1. The third column contains relative prices with the price of 1924 as

TABLE 8.1

PRICES OF TOBACCO, 1924–34; RELATIVE PRICES

Year	Price per Pound on Dec. 1, Cents	Relative Prices Price of 1924 = 100	Relative Prices Price of 1934 = 100
1924	19.0	100	86.36
1925	16 8	88	
1926	17.9	94	
1927	20.7	109	
1928	20.0	105	
1929	18.4	97	
1930	12.8	67	
1931	8.2	43	
1932	10.5	55	
1933	13 0	68	
1934	22 0	116	100

the base. The relative price for 1925 is 88, which means that the price per pound of tobacco in 1925 was 88 per cent of the price of tobacco in 1924. The relative price for 1928 was 105, which means that the 1928 price per pound of tobacco was 105 per cent of the price per pound of tobacco in 1924. The relative price is $100\left(\frac{p_1}{p_0}\right)$, where p_0 is the price in the year whose price is the base and p_1 is the price in any other year; in other words, the relative price is the quotient of the price of the commodity in year "1", and the price of the same commodity in the base year "0" multiplied by 100. These

relative prices are simple index numbers and enable one readily to compare prices of tobacco for the various years. The simple index number of price for 1930 is 67; the price index number for 1934 is 116. These relatives show the relative changes of prices of tobacco as time passes and hence form a time series. In its simplest form an index number is a name given to a term in a time series expressed as a percentage of some base, or expressed as a relative.

On examining the above relatives of prices it is seen that prices of tobacco rose from 1925 to 1927, decreased from 1927 to 1931, and increased from 1931 to 1934. These relatives or simple index numbers point out general trends of tobacco prices for these 11 years.

AGGREGATE INDEX NUMBERS

Suppose that one desires to know the relation between the level of prices in 1930 and 1934 of 10 commodities. A simple relative price as was given for each year in the last table is not adequate, as there are 10 prices for each year. The object of an index number of prices in this case is to indicate the relation of the level of the prices of these 10 commodities for the two 12-month periods. Many index numbers have been used by economists for exhibiting this relation; several index numbers will be introduced in this chapter. Table 8.2 contains prices of 10 important commodities for depression years.

TABLE 8.2

PRICES OF TEN COMMODITIES FOR 1930 TO 1934; DOLLARS

Crop	Unit	1930	1931	1932	1933	1934
Wheat.........	Bu.	$0.670	$0 390	$0.379	$0.741	$0.880
Corn..........	Bu.	.594	.321	.318	.522	.847
Oats..........	Bu.	.322	.213	.157	.334	.491
Cotton........	Lb.	.095	.057	.065	.097	.126
Sugar, B.......	Lb.	.0036	.0030	.0026	.0025	.0025
Tobacco.......	Lb.	.128	.082	.105	.130	.220
Potatoes.......	Bu.	.915	.464	.395	.823	.517
Barley.........	Bu.	.404	.325	220	.433	.710
Rye...........	Bu.	.440	.336	.276	.618	.746
Rice.....	Bu.	.784	.494	419	.778	.775
Totals.........	....	$4.3556	$2.6850	$2.3366	$4.4785	$5.3145

Price levels may be measured by comparing the aggregates of prices for various years with the sum of the prices, or the aggregate of the prices for 1930 as base. These index numbers are given in Table 8.3 for prices listed in Table 8.2; they are found by dividing the sum of the prices of a particular year by the sum of the prices of the same commodities in 1930, and multiplying by 100. The

TABLE 8.3

AGGREGATE INDEX NUMBERS FOR PRICES OF TOBACCO FROM 1930 TO 1934

Year	Index Numbers Aggregates of Prices	Index Numbers Relative Aggregates 1930 = 100	Index Numbers Relative Aggregates 1934 = 100
1930	$4 3556	100	81.96
1931	2 6850	62	
1932	2.3366	54	
1933	4.4785	103	
1934	5.3145	122	100

last number in the third column of Table 8.3, 122, is a price index number for 1934 and indicates the level of prices of 1934 compared with the level of prices of 1930. This, 122, means that the sum, or aggregate, of the prices of the 10 commodities in 1934 was 122 per cent of the aggregate of the prices for these 10 commodities in 1930. The level of prices for 1934 was much higher than that for 1930.

Let $p_0^{(i)}$ represent the price of the ith commodity for the year which is used as base and $p_1^{(i)}$ the price of this same commodity in some year "1", whose price index number is desired. The aggregative index number of prices is represented symbolically as

$$(8.1)\quad I_{ag.} = (100)\frac{\Sigma p_1^{(i)}}{\Sigma p_0^{(i)}} = \frac{p_1^{(1)}+p_1^{(2)}+p_1^{(3)}+\ldots+p_1^{(n)}}{p_0^{(1)}+p_0^{(2)}+p_0^{(3)}+\ldots+p_0^{(n)}}\times 100,$$

which is the sum of the prices for year "1" divided by the sum of the prices of the same commodities for year "0", the base year, multiplied by 100.

For example, the aggregative index number of prices for 1932 is 54, which is the sum of the prices of the 10 commodities mentioned above divided by the aggregate of the prices of these same commodities in 1930 multiplied by 100. This type of index number is called aggregative since it is the quotient of aggregates. The index numbers in Table 8.3 reveal a decided drop in the level of prices from 1930 to 1932 and a rapid rise from 1932 to 1934.

VARIOUS INDEX NUMBERS

Other index numbers of prices are obtained from relatives of prices of the commodities for various years with prices of 1930 as bases. These relatives are recorded in Table 8.4.

TABLE 8.4

Relative Prices. Prices of 1930 as Bases

Crop	Unit	1930	1931	1932	1933	1934
Wheat.......	Bu.	100	58.21	56 57	110 60	131.34
Corn..... ..	Bu.	100	54.04	53 54	87.88	142.59
Oats........	Bu.	100	66 15	48 76	103.73	152.48
Cotton......	Lb.	100	60 00	68.42	102.11	132.63
Sugar, B.....	Lb.	100	83 33	72.22	69 44	69.44
Tobacco....	Lb.	100	64.06	82.03	101.56	171.88
Potatoes.....	Bu.	100	50.71	43.17	89.95−	56.50
Rice... ..	Bu.	100	63.01	53 44	99.23	98.85+
Barley.	Bu.	100	80 45−	54.46	107.18	175.74
Rye.........	Bu.	100	76.36	62.73	140.45+	169.55
Totals.......	.....	1,000	656 32	595.34	1,012 13	1,301.00

The arithmetic average, the median, and the geometric mean of relative prices for various years have been used as index numbers of prices. These are given in Table 8.5 for data in Table 8.4.

Any one of the columns of index numbers in Table 8.5 gives a good idea of the levels of prices during these depression years. Each set of indexes reveals a decrease of the levels of prices from 1930 to 1932 and an increase from 1932 to 1934. The purpose of index numbers is to show relative changes with respect to some base.

TABLE 8.5

MEAN, MEDIAN, AND GEOMETRIC MEAN AS INDEX NUMBERS OF PRICES. PRICES OF 1930 AS BASES

Year	Arithmetic Average of Relative Prices	Median of Relative Prices	Geometric Mean of Relative Prices	Prices of 1934 = 100		
				Mean	Median	Geometric Mean
1930	100	100	100	88.35	72.72	81.76
1931	65 63	63.53	64 81			
1932	59 53	55.51	58 53			
1933	101 21	101.84	99.75			
1934	130.10	137.52	122 3	100	100	100

The arithmetic average of relative prices is

$$I_{\text{mean}} = \frac{\sum \frac{p_1}{p_0}(100)}{n}, \tag{8.2}$$

which was used in finding arithmetic averages in column 2 of Table 8.5.

Since there is an even number (10) of commodity prices the medians of relative prices were found by taking the geometric mean of the two middle relative prices for each year. The reason for taking the geometric mean of the two middle relatives instead of the arithmetic average will be given later. These medians appear in column 3 of Table 8.5.

The geometric mean of relative prices is written symbolically as

$$I_g = \sqrt[n]{\frac{p_1^{(1)}}{p_0^{(1)}} \times \frac{p_1^{(2)}}{p_0^{(2)}} \times \frac{p_1^{(3)}}{p_0^{(3)}} \times \ldots \times \frac{p_1^{(n)}}{p_0^{(n)}}}\,(100). \tag{8.3}$$

The logarithm of I_g can be written in terms of the logarithms of the relatives as

$$\log I_g = \frac{1}{n}\left[\log \frac{p_1^{(1)}}{p_0^{(1)}} + \log \frac{p_1^{(2)}}{p_0^{(2)}} + \log \frac{p_1^{(3)}}{p_0^{(3)}} + \ldots + \log \frac{p_1^{(n)}}{p_0^{(n)}}\right] + 2 = \frac{1}{n}\sum \log\left(\frac{p_1}{p_0}\right)100, \tag{8.4}$$

which is often used in finding I_g. The necessary computations for securing the geometric mean of the relative prices given in Table 8.5 for 1934 are given in Table 8.6.

TABLE 8.6

LOGARITHMS OF RELATIVE PRICES

Logarithms of Relative Prices Given in Column 7 of Table 8.4	Logarithms of Relative Prices for 1930 with Prices of 1934 as Bases
2.118 4079	1.881 5921
2.154 0970	1.845 9030
2.183 2256	1.816 7744
2.122 6469	1.877 3531
1.839 6375	2.160 3625
2.235 2127	1.764 7873
1.752 0694	1.247 9306
1.244 8768	1.755 1231
2.229 2861	1.770 7139
2.994 9956	1.005 0044
10)20.874 4555,	10)19.125 5444,
$\frac{1}{10}\Sigma \log = 2.087\ 4456$,	$\frac{1}{n}\Sigma \log = 1\ 912\ 5544$,
antilog $= I_g$,	antilog $= I_g$,
$I_g = 122.3$.	$I_g = 81.76$.

The geometric means for 1931, 1932, 1933 were obtained by finding the averages of the logarithms of relative prices given in Table 8.4 and then finding the antilogarithms of these averages. These geometric means are listed in Table 8.5 and indicate changes in the level of prices as time elapsed.

When constructing index numbers, which indicate relative changes in business conditions, the importance of certain prices should be taken into consideration. For example, the price of bread is much more important than the price of turkeys and should be given a greater weight. One way of assigning weights to various prices is to multiply prices by the respective quantities produced during the base year. In this way each commodity enters

with its price for that year and the amount sold during the base year. If quantities produced in the base year are used as weights the weighted aggregate index number of prices is

$$(8.5)\quad I_p = I_{wt} = (100)\,\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$$

$$= (100)\,\frac{p_1{}^{(1)}q_0{}^{(1)} + p_1{}^{(2)}q_0{}^{(2)} + \ldots + p_1{}^{(n)}q_0{}^{(n)}}{p_0{}^{(1)}q_0{}^{(1)} + p_0{}^{(2)}q_0{}^{(2)} + \ldots + p_0{}^{(n)}\cdot q_0{}^{(n)}}.$$

Since any year "1" may be taken as the base year, formula (8.5) becomes, if the year "1" is the base year,

$$(8.6)\qquad I_{wt} = (100)\,\frac{\Sigma p_0 q_1}{\Sigma p_1 q_1}$$

Table 8.7 contains the quantities of commodities whose prices are recorded in Table 8.2.

TABLE 8.7

PRODUCTION OF COMMODITIES, 1,000 UNITS

Crop	Unit	1930	1931	1932	1933	1934
Wheat	Bu.	889,702	932,221	745,788	528,975	469,496
Corn	Bu.	1,733,429	2,229,088	2,514,613	2,038,706	1,107,887
Oats	Bu.	1,277,379	1,126,913	1,246,548	731,500	528,815
Cotton.....	Lb.	6,966,000	8,548,000	6,501,000	6,523,500	4,865,500
Sugar B.. .	Lb.	18,398,000	15,806,000	18,140,000	22,060,000	14,962,000
Tobacco ...	Lb.	1,647,377	1,583,567	1,106,091	1,377,639	1,095,662
Potatoes....	Bu.	332,693	372,994	357,871	320,203	385,287
Rice.......	Bu	44,923	44,873	41,250	37,058	38,296
Barley	Bu.	303,752	198,543	302,042	155,825	118,929
Rye..... ..	Bu.	46,275	32,290	40,639	21,150	16,040
Totals... .	. .	31,639,530	30,874,489	30,915,842	33,794,556	23,614,885

The weighted index numbers of prices of these farm crops appear in Table 8.8 together with Fisher's* ideal index number, which is written in formula (8.7).

$$(8.7)\qquad I_{\text{ideal}} = (100)\sqrt{\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}\cdot\frac{\Sigma p_1 q_1}{\Sigma p_0 q_1}}.$$

* "The Making of Index Numbers," by Irving Fisher.

TABLE 8.8

WEIGHTED INDEX NUMBERS OF PRICES *

Year	Aggregative Index Numbers. Prices Weighted by Base Year Quantities 1930 = Base Year	Aggregative Index Numbers. Prices Weighted by Quantities of Base Year 1934	Fisher's Ideal Index Numbers
1930	100	77.51	100
1931	61 54		60 57
1932	61.03		59.70
1933	101.60		99.44
1934	138.88	100	133.86

* Prices used to the nearest penny except for sugar b. and its price was used to the nearest $\frac{1}{10}$ mill.

The numbers given in Table 8.8 reveal a decrease in the level of prices from 1930 to 1932 and an increase in the level from 1932 to 1934.

TIME REVERSAL TEST

For a number to be a good index number it should meet or satisfy the time reversal test and the factor reversal test. The time reversal test is merely a test to determine whether a given method for finding an index number will work both ways in time, backward and forward. For example, if the price level of 1934 is 150 per cent of the 1930 level, then the 1930 level of 1934 should be $66\frac{2}{3}$ per cent of the 1934 level, or 1.50 per cent times $0.66\frac{2}{3}$ per cent should be equal to unity. In other words, when price data for two years are treated by the same method and the bases are interchanged the first index number of prices should be the reciprocal of the second, and vice versa.

The simple index numbers given in Table 8.1 meet the time reversal test. The aggregative index numbers of prices also meet this test, for the index 122, for 1934, with the aggregate of prices of 1930 as bases, multiplied by the index 81.96, for 1930, with the aggregate of prices of 1934 as bases, is equal to

$$\frac{122 \times 81.96}{100 \times 100} = (9{,}999)/10{,}000 = \text{unity.}$$

In this case the base years were reversed, and the product of the index numbers expressed in percentages is unity.

The arithmetic average of relative prices does not meet this time reversal test, for

$$\frac{\sum\left(\frac{p_1}{p_0}\right)}{n} \times \frac{\sum\left(\frac{p_0}{p_1}\right)}{n}$$

is not always equal to unity. In (Table 8.5) the arithmetic average of the relative prices for 1934 with the prices of 1930 as bases is 130.1 and the arithmetic average of relative prices for 1930 with prices of 1934 as bases is 88.35. Their product divided by 10,000 is $(130.10 \cdot 88.35)/10{,}000 = 1.149$, which shows that the arithmetic average of relative prices does not always meet the time reversal test.

The median of relative prices does meet this test when there is an odd number of commodity prices; when there is an even number of commodity prices the median also meets this test if the median is considered to be the geometric mean of the two middle relative prices. This is why the geometric mean of the two middle relative prices was used for the median of relative prices instead of the arithmetic mean of the two middle relative prices in Table 8.5. The median price index, 137.52, for 1934 with respect to 1930, multiplied by the median price index, 72.72, with respect to 1934, is unity when divided by 10,000, or

$$137.52 \times 72.72/10{,}000 = 1.$$

Let $p_1^{(i)}/p_0^{(i)}$ be the middle relative price with regard to the base year "0"; this is the median of the relative prices. The ratio $p_0^{(i)}/p_1^{(i)}$ will be the median of the relative prices with respect to the base year "1". Their product is of course unity.

Let there be an even number of commodity prices and let $p_1^{(i)}/p_0^{(i)}$ and $p_1^{(i+1)}/p_0^{(i+1)}$ be the middle two relative prices with regard to year "0". The geometric mean of these two relatives is

$$\sqrt{\frac{p_1^{(i)}}{p_0^{(i)}} \cdot \frac{p_1^{(i+1)}}{p_0^{(i+1)}}};$$

this was used as the median of the relative prices. The two middle relative prices with regard to year "1" are

$$p_0^{(i)}/p_1^{(i)} \quad \text{and} \quad p_0^{(i+1)}/p_1^{(i+1)},$$

and their geometric mean is

$$\sqrt{\frac{p_0^{(i)}}{p_1^{(i)}} \cdot \frac{p_0^{(i+1)}}{p_1^{(i+1)}}}\,;$$

this is used as the median of the relatives with respect to year "1". The product of the two medians when there is an even number of commodities is unity.

The geometric mean of relative prices as an index number always meets the time reversal test for

$$\sqrt[n]{\frac{p_1^{(1)}}{p_0^{(1)}} \times \frac{p_1^{(2)}}{p_0^{(2)}} \times \frac{p_1^{(3)}}{p_0^{(3)}} \times \ldots \times \frac{p_1^{(n)}}{p_0^{(n)}}}$$

$$\times \sqrt[n]{\frac{p_0^{(1)}}{p_1^{(2)}} \times \frac{p_0^{(2)}}{p_1^{(2)}} \times \frac{p_0^{(3)}}{p_1^{(3)}} \times \ldots \times \frac{p_0^{(n)}}{p_1^{(n)}}} = 1.$$

The product of these for 1934 with respect to 1930 and for 1930 with respect to 1934 is $122.3 \times 81.76/10{,}000 = .9999 = 1$, approximately.

The weighted aggregative index numbers of prices do not always meet the first test, for the product $\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0} \times \frac{\Sigma p_0 q_1}{\Sigma p_1 q_1}$ does not always equal unity. For data in Tables 8.2 and 8.7 this product is

$$138.88 \times 77.51/10{,}000 = 1.08.$$

EXAMPLE. Prove that Fisher's ideal index number meets the time reversal test.

THE FACTOR REVERSAL TEST

To determine whether the factor reversal test is met another index number is formed by interchanging the factors in an index number. The q's are changed to p's and the p's are changed to q's; the product of the two index numbers then should give the value ratio $\frac{\Sigma p_1 q_1}{\Sigma p_0 q_0}$. The weighted aggregative index number of prices $\frac{\Sigma p_1 q_0}{\Sigma p_0 q_0}$ becomes $\frac{\Sigma q_1 p_0}{\Sigma q_0 p_0}$ when the factors are interchanged or reversed. The price index number I_p has been changed to the quantity index number I_q. Their product should be equal to the value ratio if the price index meets the factor reversal test.

The index numbers given in this chapter so far which do meet the time reversal test will be examined to determine whether or not they meet the factor reversal test.

The "ideal" index number meets this test, for

$$\sqrt{\frac{\Sigma p_1q_0}{\Sigma p_0q_0} \times \frac{\Sigma p_1q_1}{\Sigma p_0q_1}} \times \sqrt{\frac{\Sigma q_1p_0}{\Sigma q_0p_0} \times \frac{\Sigma q_1p_1}{\Sigma q_0p_1}} = \frac{\Sigma p_1q_1}{\Sigma p_0q_0};$$

that is, the index of price times the index of quantity should give the value ratio, and it does in the case of the "ideal" index number. For data in Tables 8.2 and 8.7 the "ideal" index number is 133.88 when the prices are weighted by the respective quantities and year "0" is 1930; when the factors are reversed the "ideal" index of quantities is 64.16. Their product divided by 10,000 is

$$133.86 \times 64.16/10{,}000 = .86,$$

which is equal to $\frac{\Sigma p_1q_1}{\Sigma p_0q_0}$, the value ratio. Fisher's "ideal" index number of prices meets both tests and is considered a good index number.

Consider the median of relative prices. To determine whether or not it meets the factor reversal test an index of quantities must be found by changing the p's to q's and then finding the median of the relative quantities. Let the median of the relative quantities with respect to 1930 be $\frac{q_1^{(s)}}{q_0^{(s)}}$ and the median of relative prices be $\frac{p_1^{(i)}}{p_0^{(i)}}$. Their product

$$\frac{p_1^{(i)}}{p_0^{(i)}} \times \frac{q_1^{(s)}}{q_0^{(s)}}$$

should be equal to the value ratio, $\frac{\Sigma p_1q_1}{\Sigma p_0q_0}$, if the index number here considered meets this test. The median relative price in Table 8.5 for 1934 with regard to 1930 is 137.52, and the median relative quantity of production for 1934 with regard to 1930 is 65.197. Their product divided by 10^4 is $\frac{137.52 \times 65.197}{10{,}000} = .8966$ which is not equal to the value ratio .86. Hence the median of relative prices as an index number does not meet the factor reversal test.

When the p's are changed to q's in the geometric mean of relative prices the index becomes the geometric mean of relative quantities of production, $\sqrt[n]{\frac{q_1{}^{(1)}}{q_0{}^{(1)}} \times \frac{q_1{}^{(2)}}{q_0{}^{(2)}} \times \ldots \times \frac{q_1{}^{(n)}}{q_0{}^{(n)}}}$. If the geometric mean of relative prices, as an index number, meets the factor reversal test the product

$$\sqrt[n]{\frac{p_1{}^{(1)}}{p_0{}^{(1)}} \times \frac{p_1{}^{(2)}}{p_0{}^{(2)}} \times \frac{p_1{}^{(3)}}{p_0{}^{(3)}} \times \ldots \times \frac{p_1{}^{(n)}}{p_0{}^{(n)}}} \times \sqrt[n]{\frac{q_1{}^{(1)}}{q_0{}^{(1)}} \times \frac{q_1{}^{(2)}}{q_0{}^{(2)}} \times \frac{q_1{}^{(3)}}{q_0{}^{(3)}} \times \ldots \times \frac{q_1{}^{(n)}}{q_0{}^{(n)}}}$$

should be equal to the value ratio. The geometric mean of relative prices for 1934 with respect to 1930 is 122.3, and the geometric mean of relative quantities for 1934 with regard to 1930 is 61.284. Their product divided by 10,000 is not equal to the value ratio .86; hence the geometric mean of relative prices, as an index number, does not always meet the second test.

The index number

$$I = \frac{\Sigma(q_0 + q_1)p_1}{\Sigma(q_0 + q_1)p_0}(100) \tag{8.8}$$

meets the time reversal test but does not meet the factor reversal test. Fisher recommends it as being the most practicable index number. For data in Tables 8.2 and 8.7 it is 133.86. This index nearly meets the second test. Table 8.9 allows one to compare the "ideal" index numbers of prices and the index number given by (8.8).

TABLE 8.9

IDEAL INDEX COMPARED WITH INDEX (8.8)

Year	"Ideal" Index Number of Prices; 1930 Equals Base	Index Number Given by Formula (8.8); 1930 Equals Base
1930	100	100
1931	60.57	60.52
1932	59.70	59.66
1933	99.44	99.57
1934	133.86	134.94

Values in the above table show that there is not a great difference between the "ideal" index number and that given by formula (8.8). In the light of the above discussion it seems best to use the "ideal" index number or that given by (8.8).

It would prove very helpful to have different index numbers presented in class by different members. For example the Bureau of Labor index numbers, the Federal Reserve Board Index of Factory Payrolls, Dow-Jones Stock Price Indexes, Bradstreet Monthly Commodity Index, Export Prices and Import Prices, and others.

PROBLEMS

1. Complete column 4 in Table 8.1.

2. Complete column 4 in Table 8.3.

3. Complete Table 8.5.

4. The following table gives prices of maple sirup and numbers of gallons sold from 1924 to 1934. Find the relative quantities and relative prices with the base year equal to 1924. Discuss the trends of quantities and the trends of prices.

Quantity and Price of Maple Sirup from 1924 to 1934*

Year	Quantity Sold, 1,000 Gallons	Price per Gallon
1924	3,574	$2 00
1925	2,817	2.08
1926	3,504	2.12
1927	3,429	2.05
1928	2,782	2.02
1929	2,361	2.03
1930	3,641	2.03
1931	2,213	1.72
1932	2,412	1.51
1933	2,186	1.18
1934	2,395	1 13

* These data were taken from the *U. S. Agricultural Year Book*, 1935.

5. The following table gives prices and quantities sold for 10 fruits and vegetables.

Quantity and Price of 10 Fruits and Vegetables From 1930 to 1934

	Crop	Unit	1930	1931	1932	1933	1934
Quantity	Oranges	1,000 box	55,270	50,166	51,368	47,289	58,351
Price		Box	$1.64	$1 33	$1.09	$1.59	$1.72
Quantity	Apples	1,000 bu.	102,058	106,025	85,575	74,962	75,160
Price		Bu.	$1 02	$ 65	$ 62	$.78	$.91
Quantity	Peaches	1,000 bu.	54,186	76,689	42,443	44,692	45,404
Price		Bu.	$ 88	$ 56	$ 53	$.76	$.80
Quantity	Pears	1,000 bu.	25,664	23,357	22,050	21,192	23,474
Price		Bu	$ 75	$.60	$ 39	$.55	$ 70
Quantity	G. Fruit	1,000 box	18,934	15,147	15,149	14,243	18,248
Price		Box	$1.21	$1.06	$ 84	$1 12	$.92
Quantity	Grapes	Short ton	2,443,042	1,621,315	2,203,752	1,909,581	1,775,168
Price		Ton	$19 33	$22.39	$13.16	$17 75	$20.01
Quantity	Lemons	1,000 box	7,950	7,800	6,704	7,295	7,500
Price		Box	$2 35	$1 95	$2.10	$2.35	$2.30
Quantity	Tomatoes	1,000 lb.	900,046	897,343	954,159	855,049	958,240
Price		Bu.	$1 61	$1 10	$1.03	$1.03	$1.30
Quantity	S. Pot.	1,000 bu.	53,117	63,043	78,431	65,134	67,400
Price		Bu.	$1 082	$ 725	$ 537	$.697	$ 807
Quantity	Beans	1,000 bags	13,900	12,843	10,440	12,338	10,159
Price		100-lb. bag	$4 19	$2 14	$2 01	$2.79	$3 65

Find the "ideal" index numbers for these data, with 1930 as base, and discuss the result.

6. Find the index numbers given in (8.8) for the data in problem 5.

SPLICING

Consider the following index numbers of two series.

Year	Index Number	Year	Index Number
1930	100 (base)	1934	100 (base)
1931	90	1935	103
1932	82	1936	115
1933	87	1937	120
1934	94		

If all indexes in the first series are divided by 94, the index for 1934 in the first series, we will get one series of index numbers for the entire time 1930 to 1937 with 1934 as base. These index numbers are

Year	Index Numbers
1930	106.1
1931	95.7
1932	87.2
1933	92.6
1934	100.0
1935	103.0
1936	115.0
1937	120.0

This is known as linking, or sometimes as splicing.

PROBLEMS

1. Given the following index numbers:

Year	Index Number	Year	Index Number
1926	100	1930	100
1927	104	1931	93
1928	105	1932	78
1929	124	1933	84
1930	198	1934	89
		1935	95
		1936	97
		1937	102

Combine these two series of index numbers with 1926 as base.
Combine these two series of index numbers with 1930 as base.

2. Given the following index numbers of prices:

Year	Index
1930	100
1931	90
1932	76
1933	80
1934	89
1935	95
1936	98
1937	104

Write the indexes with 1934 as base.

3. Notice index numbers in the daily or weekly newspapers.

CHAPTER 9

OBSERVATIONAL EQUATIONS

HOW OBSERVATIONAL EQUATIONS ARISE

Equations which arise from observations or direct measurements are called observational equations. Let x represent the length of an eraser and y the length of a book. Mark off on the blackboard the length of the eraser twice and the length of the book three times and then measure the entire length. This gives for the first observational equation

$$2x + 3y = 38.75 \text{ in.}$$

On the blackboard again lay off the length of the eraser and to the left of this mark lay off the length of the book twice. Measure the distance between the starting line and the finishing line; this gives another observational equation

$$x - 2y = - 10.25 \text{ in.}$$

Lay the eraser on the board five times and the book once. Measure this length. This gives a third observational equation

$$5x + y = 41.50 \text{ in.}$$

From direct observation these three equations in two unknowns arose

$$(1) \qquad 2x + 3y = 38.75,$$

$$(2) \qquad x - 2y = - 10.25,$$

$$(3) \qquad 5x + y = 41.50.$$

The question arises concerning the method of solving these equations for x and y. Three methods will be given.

First Method. Solve for x and y by using the first two equations, then by using the first and third equations, and finally by

using the second and third. Take the average of these solutions. These values are

$$\begin{array}{lll}
\text{from (1) and (2)} & x_1 = 6.68 \text{ in.}, & y_1 = 8.46 \text{ in.}, \\
\text{from (1) and (3)} & x_2 = 6.60 \text{ in.}, & y_2 = 8.52 \text{ in.}, \\
\text{from (2) and (3)} & x_3 = 6.61 \text{ in.}, & y_3 = 8.43 \text{ in.}, \\
\hline
\text{the average is} & x \ = 6.63 \text{ in.}, & y \ = 8.47 \text{ in.}
\end{array}$$

The average of these three values may be considered a solution.

When these values are substituted in the observational equations the following residual errors are obtained: 0.08, 0.06, -0.12. These errors are called residual errors because the real or true errors are not known. The sum of the squares of these residual errors is 0.0244.

Second Method. Consider that the first two observational equations are as reliable as any other two. Solving these for x and y gives $x = 6.68$ inches and $y = 8.46$ inches. When these values are used for x and y the residual errors are 0.01, 0.01, -0.36, and the sum of the squares is 0.1298. The first two residual errors would be smaller if the numbers were not rounded off to two decimal places. The sum of the squares of the residual errors for the second method is larger than that obtained from the first method.

Third Method, the Least Squares Method. Multiply the first observational equation by the coefficient of x, the second by the coefficient of x in it, and the third equation by the coefficient of x in it, and add. Do the same with the coefficients of y. These calculations give

$$\begin{array}{rcl}
4x + 6y &=& 77.50, \\
x - 2y &=& -10.25, \\
25x + 5y &=& 207.50, \\
\hline
30x + 9y &=& 274.75,
\end{array}
\qquad
\begin{array}{rcl}
6x + 9y &=& 116.25, \\
-2x + 4y &=& 20.50, \\
5x + 5y &=& 41.50, \\
\hline
9y + 14y &=& 178.25.
\end{array}$$

This gives two equations in two unknowns:

$$30x + 9y = 274.75,$$

$$9x + 14y = 178.25,$$

which are called normal equations. The following values are obtained by solving the normal equations for x and y, $x = 6.61$ inches. $y = 8.48$ inches. The residual errors are: 0.09, $+0.10$, 0.03, and the sum of the squares of the residual errors is 0.0190,

which is smaller than each of the results given by the first two methods. The third method will be considered the "best" method for solving observational equations.

By definition the "best set of values" for the unknowns will be those which make the sum of the squares of the residual error a minimum.

LEMMA. The quadratic expression $ax^2 + 2bx + c$ is a minimum if $ax + b = 0$, provided $a > 0$, and b and c are fixed real numbers.

PROOF: Multiply the quadratic by a and add and subtract b^2; this becomes

$$a^2x^2 + 2abx + b^2 + ac - b^2 = (ax + b)^2 + ac - b^2. \tag{9.1}$$

This last expression is a minimum when $ax + b = 0$, for the perfect square of a real number cannot be a negative number. When (9.1) is a minimum, so is the original quadratic expression. Hence, $ax^2 + 2bx + c$ is a minimum when $ax + b = 0$.

PROBLEMS

1. Find the minimum of the following quadratic expressions and verify your results by graphs: (a) $3x^2 - 24x + 11$; (b) $2x^2 + 7x - 5$; (c) $7x^2 + 42x + 69$.

2. Find the "best" values for x and y, the residual error, and the sum of the squares of the residual error, if the observational equations are:

$$\begin{aligned} 3x + 2y &= 23.0, \\ x - y &= -0.8, \\ -2x + 4y &= 12.3. \end{aligned}$$

LEAST SQUARES METHOD

Proof will be given showing that the third method will give the "best" values for the unknowns, or will give values for the unknowns which make the sum of the squares of the residual errors a minimum. Let the following be a set of observational equations:

$$\begin{cases} a_1x + b_1y = h_1, \\ a_2x + b_2y = h_2, \\ a_3x + b_3y = h_3, \end{cases} \tag{9.2}$$

where the a's, b's, and h's are known. These equations are as exact as the observations that were made. The normal equations, according to the third method, are

$$(9.3)\begin{cases} (a_1^2 + a_2^2 + a_3^2)x + (a_1b_1 + a_2b_2 + a_3b_3)y \\ \qquad\qquad = a_1h_1 + a_2h_2 + a_3h_3, \\ (a_1b_1 + a_2b_2 + a_3b_3)x + (b_1^2 + b_2^2 + b_3^2)y \\ \qquad\qquad = b_1h_1 + b_2h_2 + b_3h_3. \end{cases}$$

When values, which are yet unknown, for x and y are substituted in the observational equations (9.2), residual errors arise. Let e_1, e_2 and e_3 represent these errors respectively. They are

$$a_1x + b_1y - h_1 = e_1,$$
$$a_2x + b_2y - h_2 = e_2,$$
$$a_3x + b_3y - h_3 = e_3.$$

The sum of the squares of these errors is

$$e_1^2 + e_2^2 + e_3^2 = (a_1x + b_1y - h_1)^2 + (a_2x + b_2y - h_2)^2 + (a_3x + b_3y - h_3)^2,$$

or

$$\Sigma e_i^2 = (\Sigma a_i^2)x^2 + 2(\Sigma a_ib_iy - \Sigma a_ih_i)x + [(\Sigma b_i^2)y^2 - 2(\Sigma b_ih_i)y + \Sigma h_i^2].$$

This is a quadratic expression in x if y is held constant. By the lemma this quadratic expression is a minimum when

$$\Sigma a^2 \cdot x + \Sigma ab \cdot y - \Sigma ah = 0,$$

or when

$$\Sigma a^2 \cdot x + \Sigma ab \cdot y = \Sigma ah;$$

this is the first normal equation in (9.3).

Write the sum of the squares of the errors as a quadratic expression in y; this becomes

$$\Sigma e_i^2 = \Sigma b_i^2 y + 2(\Sigma abx - \Sigma bh)y + [(\Sigma a^2)x^2 - 2(\Sigma ah)x + \Sigma h^2].$$

By the lemma this quadratic expression in y is a minimum, if x is held constant, when

$$\Sigma b^2 \cdot y + \Sigma ab \cdot x - \Sigma bh = 0,$$

or when

$$\Sigma b^2 \cdot y + \Sigma ab \cdot x = \Sigma bh,$$

which is the second normal equation in (9.3).

The question arises as to what the constants are, that is, what is the constant for y when y is held constant in the quadratic expression for x and what is the constant for x when x is held constant in the quadratic expression in y. These are obtained by solving the two normal equations:

$$\left\{ \begin{aligned} \Sigma a^2 \cdot x + \Sigma ab \cdot y &= \Sigma ah, \\ \Sigma ab \cdot x + \Sigma b^2 \cdot y &= \Sigma bh, \end{aligned} \right. \tag{9.4}$$

which are the conditions which make the two quadratic expressions minimums. The solution gives values of x and y which make the sum of the squares of the residual errors a minimum, and these values by definition are the "best values" for x and y.

PROBLEMS

1. Find the best values for x and y, the residual errors, and the sum of the squares of the errors, if the observational equations are:

$$\begin{aligned} x + y &= 11.45 \text{ cm.}, \\ 3x - y &= 3.40 \text{ cm.}, \\ -x + 2y &= 11.70 \text{ cm.} \end{aligned}$$

2. Find the best values for x, y, and z, the residual errors, and the sum of the squares of the errors, if

$$\begin{aligned} 2x + y + z &= 19.2, \\ -2x + y + 2z &= 7.2, \\ -x - y - z &= 0.1, \\ 3x + 2y - z &= 11.2. \end{aligned}$$

3. Find the best values of x and y if the observational equations are:

$$\begin{aligned} 2x - y &= 0.1, \\ x + 2y &= 4.9, \end{aligned}$$

$$
\begin{aligned}
-7x - \ \ y &= +1.1, \\
3x - \ \ y &= \ \ \ 1.2, \\
-5x + 2y &= -0.9.
\end{aligned}
$$

4. If problem 3 were solved by the first method on page 145, how many pairs of simultaneous equations would have to be solved?

5. How would the lemma on page 147 be affected if $a < 0$?

6. A surveying crew measures the distance from A to B three times and the distance from B to C once. Another crew measures the distance from A to B twice and the distance from B to C three times. A third crew measures the first distance once and the second distance twice. The following observational equations were obtained: $3x + y = 33.1$ miles; $2x + 3y = 48.3$ miles; $x + 2y = 29.5$ miles. Set up the normal equations, and find the best values for the distance from A to B and the distance from B to C.

OBSERVATIONAL EQUATIONS WITH UNKNOWN CONSTANTS

In the observational equations of the preceding section the constants were known, that is, the coefficients of x, y, and z were known. In the example the eraser was laid down twice and the book three times, giving $2x + 3y = 38.75$ for the first observational equation. These coefficients arose from observations. Often problems arise which lead to observational equations in which the coefficients are unknown, while certain values of the variables or "unknowns" are given from observation.

It has been found by experience that, for fixed ages, weights and heights of men are connected by a linear relation. Let y and x represent respectively weights and heights of men of a certain age. Let the linear relation which connects weights and heights be:

$$y = a + bx. \tag{9.5}$$

The weights and heights of 10 men are found by direct measurements. When these 10 values are substituted in (9.5), 10 equations arise which will also be called observational equations. The least squares method can be used to find values of a and b, such that the sum of the squares of the residual errors shall be a minimum. The following data give the corresponding measurements of weights and heights of 10 men of a certain age:

Heights in Inches x	Weights in Pounds y	Heights in Inches x	Weights in Pounds y
66.8	139	62.9	119
66 0	117	67.6	146
70.1	150	68 6	137
68.1	166	68 4	174
64.2	122	72.3	141

The observational equations are obtained by substituting these values of x and y in the linear relation between x and y, $y = a + bx$, which will be called the predicting equation. This equation enables one to predict the weight of a man of this age when his height is known. These observational equations are:

$$
(9.6) \qquad \left\{ \begin{array}{l} \quad\quad\; x \qquad\quad\; y \\ a + 66.8\,b = 139, \\ a + 66.0\,b = 117, \\ a + 70.1\,b = 150, \\ a + 68.1\,b = 166, \\ a + 64.2\,b = 122, \\ a + 62.9\,b = 119, \\ a + 67.6\,b = 146, \\ a + 68.6\,b = 137, \\ a + 68.4\,b = 174, \\ a + 72.3\,b = 141. \end{array} \right.
$$

The normal equations are

$$
(9.7) \qquad \left\{ \begin{array}{l} Na + \Sigma x_i b = \Sigma y_i \\ \Sigma x_i a + \Sigma x_i^2 b = \Sigma x_i y_i, \end{array} \right.
$$

where i runs from 1 to N and where N is the number of measurements. In the case here considered these equations are:

$$10a + 675.00b = 1{,}411,$$

$$675.0a + 45{,}629.48b = 95{,}508,$$

from which $a = -126.460$ and $b = 3.96385$. Hence the predicting equation is

$$y = -126.460 + 3.96385x,$$

which enables one to predict weights when heights are known.

The quantity $y = a + bx = -126.46 + 3.96385x$ is the theoretical weight; the quantity y_0 is the observed value. The difference between the observed and the theoretical is called the residual error; hence residual error $= e = y_0 - y = y_0 - (a + bx) = y_0 - (-126.46 + 3.96385x)$. Substituting the weight and height of the first man in the predicting equation gives the first residual error:

$$e_1 = 139 - (-126.46 + 3.96385(66.8)) = +0.675.$$

The residual errors are:

$$\begin{aligned}
e_1 &= 139 + 126.46 - 3.96385(66.8) = +\ \ 0.675 \text{ lb.},\\
e_2 &= 117 + 126.46 - 3.96385(66.0) = -18.154 \text{ ''},\\
e_3 &= 150 + 126.46 - 3.96385(70.1) = -\ \ 1.406 \text{ ''},\\
e_4 &= 166 + 126.46 - 3.96385(68.1) = +22.522 \text{ ''},\\
e_5 &= 122 + 126.46 - 3.96385(64.2) = -\ \ 6.019 \text{ ''},\\
e_6 &= 119 + 126.46 - 3.96385(62.9) = -\ \ 3.866 \text{ ''},\\
e_7 &= 146 + 126.46 - 3.96385(67.6) = +\ \ 4.504 \text{ ''},\\
e_8 &= 137 + 126.46 - 3.96385(68.6) = -\ \ 8.460 \text{ ''},\\
e_9 &= 174 + 126.46 - 3.96385(68.4) = +29.333 \text{ ''},\\
e_{10} &= 141 + 126.46 - 3.96385(72.3) = -19.126 \text{ ''}.
\end{aligned}$$

The sum and sum of the squares of these errors are respectively:

$$\Sigma e_i = +0.003 \text{ lb.}, \qquad \Sigma e_i^2 = 2{,}208.501$$

The standard error of prediction is by definition the square root of the average of the squares of the residual errors and is

$$\sigma_e = \sqrt{\frac{\Sigma e_i^2}{N}} = \sqrt{220.8501} = 14.861 \text{ lb.}$$

Examine each of the above errors as to its size in standard errors of prediction.

The symbol σ_e has been used here; later it will be shown that this σ_e is similar to the σ used before, that is, that the predicting equation plays the role of the mean and the residual errors are actually deviations from this "mean." The sum of these residual errors will be shown to be zero. The σ_e here used is called the standard error of prediction or the standard error of estimate.

The following graph shows the straight line, $y = -126.46 + 3.96385x$, and all the residual errors as distances from this line parallel to the vertical line.

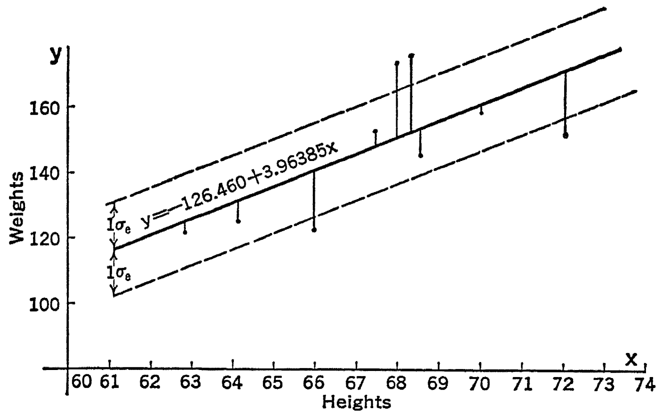


FIG. 9.0.—Predicting line for predicting weights from heights with the standard error of estimate.

PROBLEMS

1. Given the following measurements for the right thigh and the right calf of 10 male freshmen. Assume that these measurements are connected linearly. Set up the predicting equation, the observational equations, and the normal equations, and find the predicting equation for predicting right thigh measurements from right calf measurements. Find the sum of the residual errors and the standard error of prediction.

RIGHT THIGH	RIGHT CALF	RIGHT THIGH	RIGHT CALF
19.0	12.8	22.7	14.7
18.6	11.7	20.9	13.1
21.0	14.4	22.1	14.2
20.8	13.5	24.0	15.5
18.0	12.5	17.5	12.2

Plot the predicting equation and show the errors in the plot.

2. In what range on the horizontal axis does the predicting line have meaning?

3. Given a set of items; what is the sum of the deviations of the items from the mean?

4. How far from zero is the sum of the residual errors in problem 1? What does this suggest about the predicting equation?

5. On each side of the predicting line draw a line parallel to it at a distance of 1 standard error of prediction away. Count how many of the ordinates for right thigh measurements fall in this band.

6. Given a set of variates distributed normally. What percentage of the variates falls within 1 standard deviation from the mean of these variates?

7. What percentage of the measurements falls within the band mentioned in problem 5?

SOLUTION OF THE NORMAL EQUATIONS AND THE STANDARD ERROR OF PREDICTION IN TERMS OF FUNDAMENTAL SUMMATIONS

Let the relation between x and y be linear, that is, $y = a + bx$, and let the following be a set of n observational equations, where a and b are unknowns and the x's and y's arise from measurements:

$$\begin{aligned} a + bx_1 &= y_1, \\ a + bx_2 &= y_2, \\ a + bx_3 &= y_3, \\ &\cdots\cdots \\ a + bx_n &= y_n, \end{aligned}$$

The normal equations are

$$\begin{aligned} na + \Sigma x \cdot b &= \Sigma y, \\ \Sigma x \cdot a + \Sigma x^2 \cdot b &= \Sigma xy, \end{aligned}$$

from which

$$a = \frac{\Sigma y\, \Sigma x^2 - \Sigma x\, \Sigma xy}{n\, \Sigma x^2 - (\Sigma x)^2}, \quad \text{and} \quad b = \frac{n\, \Sigma xy - \Sigma x\, \Sigma y}{n\, \Sigma x^2 - (\Sigma x)^2}.$$

These express the constants a and b in terms of the fundamental summations, n, Σx, Σx^2, Σy, Σxy.

The standard error of prediction can also be expressed in terms of these fundamental summations together with Σy^2, as:

$$\begin{aligned} e_1{}^2 &= [y_1 - (a+bx_1)]^2 = y_1{}^2 + a^2 + b^2x_1{}^2 + 2abx_1 - 2ay_1 - 2bx_1y_1, \\ e_2{}^2 &= [y_2 - (a+bx_2)]^2 = y_2{}^2 + a^2 + b^2x_2{}^2 + 2abx_2 - 2ay_2 - 2bx_2y_2, \\ e_3{}^2 &= [y_3 - (a+bx_3)]^2 = y_3{}^2 + a^2 + b^2x_3{}^2 + 2abx_3 - 2ay_3 - 2bx_3y_3, \\ &\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ e_n{}^2 &= [y_n - (a+bx_n)]^2 = y_n{}^2 + a^2 + b^2x_n{}^2 + 2abx_n - 2ay_n - 2bx_ny_n. \end{aligned}$$

$$\Sigma e^2 = \Sigma y^2 + [na^2 + \Sigma x^2 \cdot b^2 + 2ab \cdot \Sigma x] - 2a \cdot \Sigma y - 2b\Sigma xy. \tag{9.8}$$

Multiply the first normal equation by a and the second by b, which gives on adding:

$$na^2 + 2ab \cdot \Sigma x + b^2 \cdot \Sigma x^2 = a \cdot \Sigma y + b \cdot \Sigma xy.$$

Substituting this expression in (9.8) for the quantity in brackets gives for the sum of the squares of the residual errors:

$$\text{(9.9)} \qquad \begin{aligned} \Sigma e^2 &= \Sigma y^2 + a\,\Sigma y + b\,\Sigma xy - 2a\,\Sigma y - 2b\,\Sigma xy \\ &= \Sigma y^2 - a \cdot \Sigma y - b \cdot \Sigma xy. \end{aligned}$$

Then

$$\text{(9.10)} \qquad \sigma_e = \sqrt{\frac{\Sigma e^2}{n}} = \sqrt{\frac{\Sigma y^2 - a \cdot \Sigma y - b\,\Sigma \cdot xy}{n}},$$

which expresses the standard error of prediction in terms of the fundamental summations, Σx, Σy, Σxy, Σx^2, Σy^2, and n.

When these six summations are known, the predicting equation can be written with little trouble; the standard error of prediction can also be written at once. It is not necessary to write the observational equations, or even the normal equations, if the values of a and b on page 154 are used.

PROBLEMS

1. Given the following chest and waist measurements of men 18 years of age. Assume that chest measurements are connected linearly with waist measurements. Find the predicting equation for predicting chest measurements from waist measurements. Find the standard error of prediction.

Chest Measurement	Waist Measurement	Chest Measurement	Waist Measurement
32.0	28.0	31.0	25.6
37.0	34.5	32.2	30.0
31.4	27.0	36.1	32.1
36.5	35.8	29.0	25.0
29.5	26.0	31.0	29.0

Plot the data and draw the line of prediction.

2. The standard error of prediction is 14.2 cubic feet, and the sum of the squares of the residual errors is 30,850.92. Find the value of n, the number of observations which were used to find σ_e.

PREDICTING EQUATION IN TERMS OF DEVIATIONS FROM THE MEAN

From earlier material it is known that

$$\Sigma\bar{x}^{2*} = \Sigma(x - M_x)^2 = \Sigma x^2 - (\Sigma x)^2/n \tag{9.11}$$

and

$$\Sigma x^2 = \Sigma\bar{x}^2 + (\Sigma x)^2/n;$$

also

$$\Sigma\bar{y}^2 = \Sigma y^2 - (\Sigma y)^2/n \quad \text{and} \quad \Sigma y^2 = \Sigma\bar{y}^2 + (\Sigma y)^2/n. \tag{9.12}$$

A similar formula will be derived for $\Sigma\bar{x}\bar{y}$ as follows:

$$\begin{aligned}\Sigma\bar{x}\bar{y} = \Sigma(x - M_x)(y - M_y) &= \Sigma(xy - xM_y - yM_x + M_xM_y)\\ &= \Sigma xy - M_y\cdot\Sigma x - M_x\cdot\Sigma y + nM_xM_y\\ &= \Sigma xy - \Sigma x\cdot\Sigma y/n - \Sigma y\cdot\Sigma x/n + (\Sigma x)(\Sigma y)/n\end{aligned}$$

or

$$\Sigma\bar{x}\bar{y} = \Sigma xy - \frac{\Sigma x\Sigma y}{n}, \tag{9.13}$$

and

$$\Sigma xy = \Sigma\bar{x}\bar{y} + \frac{\Sigma x\cdot\Sigma y}{n}.$$

Substituting the values of Σxy and Σx^2 from the above in the value of b given on page 154 gives:

$$b = \frac{n(\Sigma\bar{x}\bar{y} + \Sigma x\Sigma y/n) - \Sigma x\Sigma y}{n[\Sigma\bar{x}^2 + (\Sigma x)^2/n] - (\Sigma x)^2} = \frac{\Sigma\bar{x}\bar{y}}{\Sigma\bar{x}^2};$$

The value of a found by solving the first normal equation is

$$a = M_y - bM_x.$$

The predicting equation is therefore

$$y = M_y - M_x(\Sigma\bar{x}\bar{y}/\Sigma\bar{x}^2) + (\Sigma\bar{x}\bar{y}/\Sigma\bar{x}^2)x. \tag{9.14}$$

The standard error of prediction reduces to

$$\sigma_e = \sqrt{\frac{\Sigma\bar{y}^2}{n} - \frac{(\Sigma\bar{x}\bar{y})^2}{n\Sigma\bar{x}^2}}. \tag{9.15}$$

* Note that $\bar{x}$ (read bar x) and $\bar{y}$ are derivations of x and y from their respective means.

EXAMPLE. The following data are measurements of 10 men's weights and chest measurements.

y Wts. Lb.	x Ch. Meas. In.	y Wts. Lb.	x Ch. Meas. In.
139	34 5	119	34.5
117	34.5	146	37.0
150	37 6	137	33.0
166	38 0	174	38.3
122	33.4	141	34.0

It is assumed that weights and chest measurements are connected linearly by the equation $y = a + bx$. The predicting equation and standard error of prediction will be found by the new formulas.

According to (9.11), (9.12), and (9.13)

$$\Sigma\bar{x}^2 = \Sigma x^2 - (\Sigma x)^2/n = 12{,}624.96 - 12{,}588.30 = 36.66,$$

$$\Sigma\bar{y}^2 = \Sigma y^2 - (\Sigma y)^2/n = 202{,}353 - 199{,}092.1 = 3.260.9,$$

$$\Sigma\bar{x}\bar{y} = \Sigma xy - (\Sigma x)\,(\Sigma y)/n = 50{,}341.5 - 50{,}062.3 = 279.2,$$

$$b = \Sigma\bar{x}\bar{y}/\Sigma\bar{x}^2 = 279.2/36.66 = 7.61593,$$

$$a = M_y - M_x b = 141.1 - 35.48(7.61593) = -\,129.113.$$

The predicting equation is therefore

$$y = -\,129.113 + 7.61593\,x.$$

The standard error of prediction is $\sigma_e = 10.65^+$, which is smaller than the standard error of prediction found on page 152 when weights were predicted from heights. The smaller σ_e shows that, for these 10 men, chest measurement is better for predicting weight than height. The illustrations on page 152 and on page 157 point out the importance of the standard error of prediction.

The above results can be derived by easier methods. Let the predicting equation be written in terms of deviations from the means, that is

$$\bar{y} = a' + b\bar{x}. \tag{9.16}$$

The normal equations which are obtained from the observational equations are:

$$na' + \Sigma\bar{x}b = \Sigma\bar{y},$$

$$\Sigma\bar{x}a' + \Sigma\bar{x}^2b = \Sigma\bar{x}\bar{y}.$$

The $\Sigma\bar{x} = 0$ and $\Sigma\bar{y} = 0$, hence the first normal equation gives $a' = 0$ and the second normal equation gives

$$b = \Sigma\bar{x}\bar{y}/\Sigma\bar{x}^2,$$

which is the same as b found on page 156. The predicting equation becomes

$$(9.17) \qquad \bar{y} = b\bar{x}, \quad \text{or} \quad \bar{y} = \frac{\Sigma\bar{x}\bar{y}}{\Sigma\bar{x}^2} \cdot \bar{x},$$

or

$$y - M_y = \frac{\Sigma\bar{x}\bar{y}}{\Sigma\bar{x}^2}(x - M_x),$$

or

$$y = M_y - (\Sigma\bar{x}\bar{y}/\Sigma\bar{x}^2)M_x + (\Sigma\bar{x}\bar{y}/\Sigma\bar{x}^2)x,$$

which is the same equation as found in (9.14). Equation (9.16) is the same as equation (9.14).

Consider the predicting equation in the form $y = a + bx$. Translate the origin to the point (M_x, M_y) by the transformation equations used in analytical geometry.

$$y = y' + M_y,\ y' = y - M_y = \bar{y}, \text{ and } x = x' + M_x.$$

These used in the above predicting equation give:

$$y' + M_y = a + b(x' + M_x), \quad \text{or} \quad \bar{y} = a - M_y + bM_x + b\bar{x},$$

which becomes, if $a = M_y - bM_x$,

$$\bar{y} = b\bar{x},$$

as equation (9.17).

Substitute the coordinates of the point (M_x, M_y) in the predicting equation (9.14). These coordinates satisfy this equation, hence the point (M_x, M_y) lies on the predicting line. Therefore the a found here is the same as the a found before; hence line (9.14) is the same as line (9.17).

The graph shows the predicting equation plotted with respect to the two sets of axes.

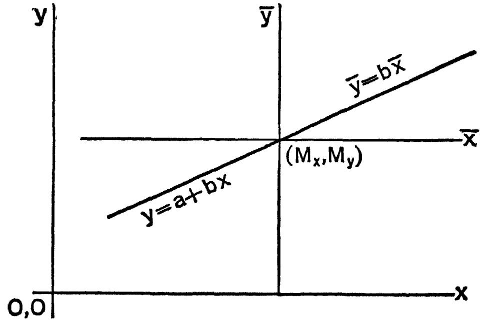


FIG. 9.1.—Equation of the predicting line in terms of original units and deviations from the means.

PROBLEMS

1. Given the figures below for school census for a certain town from 1917 to 1936. Find $\Sigma\bar{x}^2$, $\Sigma\bar{x}\bar{y}$ from formulas (9.11) and (9.13), then a and b from formulas on page 156, and the linear predicting equation from (9.14) for predicting census figures from years. Find σ_e from (9.15).

Year	Census	Year	Census
1917	1,840	1927	2,510
1918	1,850	1928	2,600
1919	1,931	1929	2,690
1920	2,003	1930	2,782
1921	2,048	1931	2,705
1922	2,054	1932	2,820
1923	2,183	1933	2,910
1924	2,340	1934	3,045
1925	2,387	1935	3,183
1926	2,460	1936	3,300

2. Derive formula (9.15).

3. About what is the probability that one will predict a value within 1 σ_e of the predicting line?

4. Does the predicting equation give exact values?

5. Using the predicting equation found in problem 1 predict the census for 1920 and for 1936. How many standard errors of prediction did you miss the observed census figures?

COMPUTATIONS SHORTENED BY USE OF THE PROVISIONAL MEAN

The constant b and the standard error of prediction have been developed in terms of the following summations: n, $\Sigma\bar{x}^2$, $\Sigma\bar{y}^2$, $\Sigma\bar{x}\bar{y}$. These can be computed by subtracting provisional means from the variates of the original observations. Let the deviation from a provisional mean for the x's be represented by

$$x'_i = x_i - h,$$

where h is the provisional mean. The deviation of x'_i from its mean is equal to the corresponding deviation of x from its mean as shown in an earlier chapter. Hence the deviations from the mean in the new set of variates are equal respectively to the deviations of the variates in the original set from their mean. If values of the original observations are large numbers they can be made smaller by subtracting a provisional mean from each variate. The

quantities a, b, and σ_e can now be written in terms of the new summations:

$$n,\ \Sigma x',\ \Sigma y',\ \Sigma x'^2,\ \Sigma y'^2,\ \Sigma x'y',$$

as

$$(9.18)\qquad \begin{cases} b = \dfrac{\Sigma\bar{x}\bar{y}}{\Sigma\bar{x}} = \dfrac{\Sigma\bar{x}'\bar{y}'}{\Sigma\bar{x}'^2} = \dfrac{\Sigma x'y' - (\Sigma x')(\Sigma y')/n}{\Sigma x'^2 - (\Sigma x')^2/n}, \\ a = M_y - M_x b = M_{y'} + h_y - (M_{x'} + h_x)b, \end{cases}$$

where h_y is the provisional mean taken from each y and h_x is the provisional mean taken from each x measurement.

The standard error of prediction can now be written in terms of the new set of data as follows:

$$(9.19)\quad \sigma_e = \sqrt{\frac{\Sigma\bar{y}^2}{n} - \frac{(\Sigma\bar{x}\bar{y})^2}{n\Sigma\bar{x}^2}} = \sqrt{\frac{\Sigma\bar{y}'^2}{n} - \frac{(\Sigma\bar{x}'\bar{y}')^2}{n\Sigma\bar{x}'^2}}$$

$$= \sqrt{\frac{\Sigma y'^2 - (\Sigma y')^2/n}{n} - \frac{[\Sigma x'y' - (\Sigma x')(\Sigma y')/n]^2}{n[\Sigma x'^2 - (\Sigma x')^2/n]}}.$$

These formulas appear to be more complicated than the others, but they are not so complicated as they appear. In analyzing a problem of this nature the following 6 summations must be computed:

$$n,\ \Sigma x',\ \Sigma x'^2,\ \Sigma y',\ \Sigma y'^2,\ \Sigma x'y'.$$

The values of a, b, and σ_e can be immediately obtained. It is best to calculate the following

$$(9.20)\qquad \Sigma\bar{y}'^2 = \Sigma y'^2 - \frac{(\Sigma y')^2}{n};\ \Sigma\bar{x}'^2 = \Sigma x'^2 - \frac{(\Sigma x')^2}{n},$$

$$\Sigma\bar{x}'\bar{y}' = \Sigma x'y' - \frac{(\Sigma x')(\Sigma y')}{n},$$

and then substitute them in (9.19) and the formulas for a and b.

Example. Given the following weights and corresponding right thigh measurements of the 10 men of the example at the beginning of this chapter. Assume that weights and right thigh measurements are related linearly; find the predicting equation and σ_e.

y Weights in Pounds	x Right Thigh Meas. in Inches	y Weights in Pounds	x Right Thigh Meas. in Inches
139	20.0	119	19.5
117	19.0	146	22.2
150	20.4	137	21.5
166	24.0	174	24.2
122	19.5	141	21.2

Let 140 be h_y and $21 = h_x$. After these have been subtracted from the above data we get the following

y'	x'	y'	x'
-1	-1.0	-21	-1.5
-23	$-2\ 0$	$+6$	$+1.2$
$+10$	$-0\ 6$	-3	$+0.5$
$+26$	$+3.0$	$+34$	$+3.2$
-18	-1.5	$+1$	$+0.2$

From these the fundamental summations are found to be

$$\Sigma x' = 1.5,\ \Sigma y' = 11,\ \Sigma x'^2 = 30.83,\ \Sigma y'^2 = 3{,}273$$

$$\Sigma x'y' = 292.2,\ \Sigma \bar{x}'^2 = 30.605,\ \Sigma \bar{y}'^2 = 3{,}260.9,\ \Sigma \bar{x}'\bar{y}' = 290.55,$$

from which $a = -59.689$, $b = 9.49355$, $\sigma_e = 7.0887$. The predicting equation is

$$y = -59.6885 + 9.4935\,x.$$

The standard deviation, $\sigma_e = 7.089$ pounds, is much smaller than that found when predicting weights from heights and weights from chest measurements. Hence right thigh measurements, for these 10 men, are better for predicting weights than heights or chest measurements.

Theorem 9.1. The sum of the residual errors is equal to zero.

Proof: Let us use the predicting line in the form $\bar{y} = b\bar{x}$. The residual errors are now

$$\begin{aligned} e_1 &= \bar{y}_1 - b\bar{x}_1, \\ e_2 &= \bar{y}_2 - b\bar{x}_2, \\ e_3 &= \bar{y}_3 - b\bar{x}_3, \\ &\cdots\cdots\cdots\cdots \\ e_n &= \bar{y}_n - b\bar{x}_n \end{aligned}$$

$$\text{Add} \quad \Sigma e = \Sigma \bar{y} - b\Sigma \bar{x} = 0 + b\cdot 0 = 0.$$

This shows that the predicting line is similar to a real mean, since the sum of the deviations from it is zero. The standard error of prediction is very similar to a standard deviation. Sometimes the predicting line is spoken of as a "moving average."

PROBLEMS

1. The following table contains the average length of intestines of birds in centimeters and the average weight of the body in grams. Find the (linear) predicting equation for predicting intestine length from body weight and the standard error of prediction.*

Average Length of Intestines of Several Birds, Centimeters	Average Weight of Several Birds, Grams
4.3	1.5
5.8	2.7
6 5	3.6
7.3	4 2
8.4	5.4
9.0	5.9
9.7	6.5
10.2	7.3
11.0	8.1
11.6	8.8
12.4	9.7
12.6	9.8

2. The following data give the ages in years and heights in inches for a group of girls, where x represents ages and y represents heights:

x	y	x	y	x	y	x	y	x	y	x	y
4	40	7	50	9	56	11	50	13	60	16	60
4	42	7	50	9	58	11	54	13	60	16	64
5	42	7	52	9	58	11	54	13	58	16	62
5	44	8	54	10	48	11	58	13	64	16	62
5	46	8	54	10	46	11	60	14	58	17	64
6	44	8	56	10	50	11	60	14	66	17	66
6	46	8	56	10	50	12	56	14	64	17	64
6	50	8	58	10	54	12	58	14	60	18	64
6	48	9	50	10	56	12	58	15	66	18	66
7	48	9	52	10	60	12	64	15	64	18	62
7	48	9	54	11	48	12	62	15	62	19	68

* Taken from "Experiments and the Digestion of Food by Birds," by James Stevenson, *Wilson Bulletin*, Vol. XLV., No. 4, 1935, pages 155–167.

Find the predicting equation for predicting heights from ages. Find the average height for each group, and plot these points on the same graph with the predicting line. What does the predicting line actually give for a predicted value?

PREDICTING EQUATIONS IN MORE THAN TWO UNKNOWNS

Assume that 3 variables x, y and z are connected linearly as

$$y = a + bx + cz, \tag{9.21}$$

where y is predicted from the values of x and z. Here there will be n observational equations and three normal equations in three unknowns. The number of normal equations can be reduced to two if the predicting equation is written in terms of the deviations from the means, viz.:

$$\bar{y} = a' + b\bar{x} + c\bar{z}.$$

There are still 3 unknowns, a', b, and c, and it does not appear as though there will be a reduction of normal equations. Examine the normal equations

$$na' + \Sigma\bar{x}b + \Sigma\bar{z}c = \Sigma\bar{y},$$

$$\Sigma\bar{x}\cdot a' + \Sigma\bar{x}^2\cdot b + \Sigma\bar{x}\bar{z}\cdot c = \Sigma\bar{x}\bar{y},$$

$$\Sigma\bar{z}\cdot a' + \Sigma\bar{x}\bar{z}\cdot b + \Sigma\bar{z}^2\cdot c = \Sigma\bar{z}\bar{y},$$

The sum of the deviations from the mean is zero; hence $\Sigma\bar{x}, = \Sigma\bar{y} = \Sigma\bar{z} = 0$, and the first normal equation reduces to $na' = 0$; hence $a' = 0$. The above three normal equations become:

$$\Sigma\bar{x}^2\cdot b + \Sigma\bar{x}\bar{z}\cdot c = \Sigma\bar{x}\bar{y},$$

$$\Sigma\bar{x}\bar{z}\cdot b + \Sigma\bar{z}^2\cdot c = \Sigma\bar{z}\bar{y}.$$

The predicting equation becomes

$$\bar{y} = b\bar{x} + c\bar{z}. \tag{9.22}$$

To be able to use this predicting equation for predicting, the deviations from the mean of the x's and the deviations from the mean of the z's must be known; this then gives for the predicted value of the y a deviation from the mean of the y's. Measurements, as a rule, are not given in terms of deviations from the mean; hence the last predicting equation is impracticable. Predicting equation (9.21) should be used for predicting after the

constants are found by using (9.22). The quantities b and c in (9.22) are the same as in (9.21), for $\bar{y} = b\bar{x} + c\bar{z}$ can be written as

$$y - M_y = b(x - M_x) + c(z - M_z), \text{ or } y = (M_y - bM_x - cM_z) + bx + cz,$$

which is (9.21) if

$$a = M_y - bM_x - cM_z. \tag{9.23}$$

Equation (9.22) is used to reduce the number of normal equations by one. This enables one to find b and c much more quickly. When b and c are found, a can be found easily by using (9.23).

Equation (9.22) can be obtained from (9.21), as was done on page 158.

The standard error of prediction is for this case:

$$\sigma_e = \sqrt{\frac{\Sigma\bar{y}^2 - b\cdot\Sigma\bar{x}\bar{y} - c\cdot\Sigma\bar{z}\bar{y}}{n}}. \tag{9.24}$$

Consider the problem of finding the predicting equation for predicting weights from heights and thigh measurements by using the data on pages 151 and 161. Let the linear relation be

$$y = a + bx + cz, \text{ or } \bar{y} = b\bar{x} + c\bar{z},$$

where y represents weights, x represents heights, and z right thigh measurements. The normal equations, if the second equation is used, are

$$\Sigma\bar{x}^2\cdot b + \Sigma\bar{x}\bar{z}\cdot c = \Sigma\bar{x}\bar{y},$$

$$\Sigma\bar{z}\bar{x}\cdot b + \Sigma\bar{z}^2\cdot c = \Sigma\bar{z}\bar{y}.$$

or

$$\Sigma\bar{x}'^2 b + \Sigma\bar{x}'\bar{z}'c = \Sigma\bar{x}'\bar{y}',$$

$$\Sigma\bar{x}'\bar{z}'b + \Sigma\bar{z}'^2 c = \Sigma\bar{z}'\bar{y}'.$$

Using $h_y = 140$, $h_x = 67$, $h_z = 21$, the data reduce to the following:

y'	x'	z'	y'	x'	z'
− 1	−0.2	−1.0	−21	−4.1	−1.5
−23	−1.0	−2.0	+ 6	+0 6	+1.2
+10	+3.1	−0.6	− 3	+1.6	+0.5
+26	+1.1	+3.0	+34	+1.4	+3.2
−18	−2.8	−1.5	+ 1	+5.3	+0.2

$$\Sigma y' = 11, \Sigma x' = 5.0, \Sigma z' = 1.5,$$

$$\Sigma y'^2 = 3{,}273, \Sigma x'^2 = 69.48, \Sigma z'^2 = 30.83,$$

$$\Sigma\bar{y}'^2 = 3{,}260.9,\ \Sigma\bar{x}'^2 = 66.98,\ \Sigma\bar{z}'^2 = 30.605,$$

$$\Sigma x'y' = 271.0,\ \Sigma x'z' = 21.05,\ \Sigma z'y' = 292.2,$$

$$\Sigma\bar{x}'\bar{y}' = 265.5,\ \Sigma\bar{x}'\bar{z}' = 20.30,\ \Sigma\bar{z}'\bar{y}' = 290.55.$$

The normal equations are

$$66.98b + 20.3c \quad = 265.5,$$

$$20.30b + 30.605c = 290.5.$$

from which

$$b = 1.362, \quad c = 8.586.$$

From (9.23)

$$a = -132.429.$$

The predicting equation is

$$y = -132.429 + 1.362x + 8.586z,$$

and the standard error of prediction is $\sigma_e = 6.361$ pounds.

The size of the standard error of prediction shows that weights are better predicted from heights and right thigh measurements together than from either one. Compare the σ_e's on pages 157 and 161.

PROBLEMS

1. If $y = a + bx + cz$ and there are n measurements for each variable, set up the observational equations and the normal equations for finding of a, b, and c.

2. Assume that weights, heights, and chest measurements for men are connected linearly. Find the predicting equation for weights by using data on pages 151 and 157. Find σ_e. Compare with the above illustration.

3. Assume a linear relation between weight, chest measurements, and right thigh measurements. Find the predicting equation for weights by using data on pages 151, 157, and 161.

4. Compare the sizes of the σ_e found in problems 2 and 3 and on page 165, and determine which two measurements are best for predicting weights.

5. Derive formula (9.24).

6. Prove by translating the axes that the predicting plane $y = a + bx + cz$ is the same as the predicting plane $\bar{y} = b\bar{x} + c\bar{z}$.

7. Prove that the point (M_x, M_y, M_z) is on the predicting plane $y = a + bx + cz$.

SOLUTION OF SIMULTANEOUS EQUATIONS

Given the following equations in three unknowns, where the coefficients of the unknowns are large numbers:

$$(1) \qquad 10.0\,a + 354.80\,b + 211.50\,c = 1{,}411.00,$$

$$(2) \qquad 354.8\,a + 12{,}624.96\,b + 7{,}527.15\,c = 5{,}034.15,$$

$$(3) \qquad 211.5\,a + 7{,}527.15\,b + 4{,}053.83\,c = 30{,}133.20.$$

Divide each by the coefficient of a, getting:

$$(1) \qquad a + 35.4800\,b + 21.1500\,c = 141.100,$$

$$(2) \qquad a + 35.5833\,b + 21.2152\,c = 141.887,$$

$$(3) \qquad a + 35.5894\,b + 21.2947\,c = 142.474.$$

Subtract (2) from (3) and (1) from (2), and then multiply by 10,000.

This gives $\qquad 61\,b + 795\,c = 5{,}870,$

$$1{,}094\,b + 1{,}447\,c = 13{,}740.$$

Divide by the coefficients of b, getting

$$b + 13.033\,c = 96.230,$$

$$b + 1.323\,c = 12.559.$$

Subtracting we get, $\qquad 11.710\,c = 83.671,$

Therefore $\quad c = 7.1671, \quad b = 3.051, \quad a = -119.154.$

It is best to subtract the first two equations, then the next two, and then the next two, etc., in preference to using one as the subtrahend throughout, for this one subtrahend might have been more incorrect than the others, and hence more errors would creep into the results. By using the first two, the second two, etc., there will be a tendency to equalize the errors in the different equations or make all enter in about the same way or to the same weight.

This method of reducing the coefficients of the first unknown to unity is far easier than solving by determinants or by other algebraic methods, especially when the coefficients are large.

PROBLEMS

1. Assume that weights, heights, chest measurements, and right thigh measurements are connected linearly. Find the predicting equation for weights, using data on pages 151, 157, and 161. Find σ_e.

2. Use data in problem 1 and those on pages 152, 157, 161, and 165. Do heights, chest measurements, and right thigh measurements enable one to predict weights better than heights alone? Chest measurements? Right thigh measurements? Better than heights and chest measurements, or heights and thigh measurements or chest and thigh measurements? What is the determining factor?

3. Using the predicting equation in problem 1 find the predicted weight of the third individual; the weight of the tenth man. How large are the errors in terms of σ_e?

4. Prove that, if $y = a + bx + cz + du + gw$, then

$$\sigma_e = \sqrt{\frac{\Sigma y^2 - a\Sigma y - b\Sigma xy - c\Sigma zy - d\Sigma uy - g\Sigma wy}{n}}$$

and that

$$\sigma_e = \sqrt{\frac{\Sigma \bar{y}^2 - b\Sigma \bar{x}\bar{y} - c\Sigma \bar{z}\bar{y} - d\Sigma \bar{u}\bar{y} - g\Sigma \bar{w}\bar{y}}{n}}$$

5. Prove that, if the predicting equation is $y = a + bx$, then

$$\sigma_e = \sigma_y \sqrt{1 - \frac{(\Sigma \bar{x}\bar{y})^2}{\Sigma \bar{x}^2 \Sigma \bar{y}^2}}.$$

6. When will the standard error of prediction be equal to zero?

7. Discuss the situation when the denominator in a and b on page 154 is equal to zero.

8. Set up the normal equations when the predicting equation is $y = a + bx + cx^2$, if there are given n corresponding measurements for x and y.

9. Set up the normal equations when y and x are connected by the relation $y = a/x$, if n corresponding measurements are given for x and y.

10. Do the values of a and b, found on page 160, depend on the provisional means h_x and h_y?

CHAPTER 10

CORRELATION COEFFICIENT

LINEAR CORRELATION

Assume that two variables x and y are connected linearly by

$$y = a + bx, \tag{10.1}$$

where the x's and y's are given from observation or from direct measurements. When these values are substituted in this equation, n observational equations arise, from which the following normal equations are obtained:

$$n \cdot a + \Sigma x \cdot b = \Sigma y,$$

$$\Sigma x \cdot a + \Sigma x^2 \cdot b = \Sigma xy.$$

The solution of these normal equations gives

$$a = M_y - M_x \frac{\Sigma \bar{x}\bar{y}}{\Sigma \bar{x}^2}, \quad b = \frac{\Sigma \bar{x}\bar{y}}{\Sigma \bar{x}^2}.$$

The standard error of prediction is found to be

$$\sigma_e = \sqrt{\frac{\Sigma \bar{y}^2}{n} - \frac{(\Sigma \bar{x}\bar{y})^2}{n\Sigma \bar{x}^2}} = \sigma_y \sqrt{1 - \frac{(\Sigma \bar{x}\bar{y})^2}{(\Sigma \bar{x}^2)(\Sigma \bar{y}^2)}}, \tag{10.2}$$

or

$$\sigma_e = \sigma_y \sqrt{1 - r_{yx}^2}, \tag{10.3}$$

where

$$r_{yx} = \frac{\Sigma \bar{x}\bar{y}}{\sqrt{\Sigma \bar{x}^2 \Sigma \bar{y}^2}}, \tag{10.4}$$

which is called the Pearson linear correlation coefficient between x and y, or the linear correlation coefficient, which may be written

$$r_{yx} = \frac{\Sigma \bar{x}\bar{y}}{n\sigma_x \cdot \sigma_y}. \tag{10.5}$$

The quantity r_{yx} can never be greater than 1 and it can never be less than -1, for if $|r_{yx}| > 1$, the standard error of prediction which is positive or zero would be imaginary, as is seen by examining (10.3).

If $r = \pm 1$, then $\sigma_e = 0$, and hence all the residual errors are equal to zero and hence all the values of y will fall on the line of prediction; this means perfect correlation between x and y.

If $r = 0$, then $\sigma_e = \sigma_y$. To be able to understand what this means let us write the predicting equation in terms of the correlation coefficient, and the quantities σ_x, σ_y, M_x, M_y. This is

$$y = M_y - M_x \cdot \frac{\Sigma \bar{x}\bar{y}}{\Sigma \bar{x}^2} + \frac{\Sigma \bar{x}\bar{y}}{\Sigma \bar{x}^2} \cdot x. \quad \text{[See (9.14).]}$$

From (10.5)

$$\Sigma \bar{x}\bar{y} = n\sigma_x \cdot \sigma_y \cdot r_{xy}.$$

Therefore the predicting equation is

$$(10.6) \qquad y = M_y - M_x \cdot \frac{\sigma_y}{\sigma_x} \cdot r_{yx} + r_{yx} \cdot \frac{\sigma_y}{\sigma_x} \cdot x,$$

which gives the predicting equation in terms of the means, standard deviations, and correlation coefficient; hence, if $r_{yx} = 0$, the predicting equation becomes

$$(10.7) \qquad y = M_y,$$

which shows that for any value of the variable x the best value of y will always be the mean of the y's. For example, if heights of several people are known, then the predicted weights for each individual would be the average of all weights, that is, if the correlation coefficient between heights and weight were zero. This average of all weights, M_y, would represent the weight of a heavy man as well as a light man. This is the same as saying that for any height there is one and only one weight, which is the average of all weights. Here the predicting curve is a straight line parallel to the horizontal axis. This means that there is no relation between x and y, that is, y does not depend upon x.

If $r = -1$, y becomes larger when x becomes smaller, as can be seen from predicting equation (10.6). Correlation here is perfect. If two test scores were connected in such a way as to yield a linear correlation coefficient of -1, then high scores in one test would correspond to low scores in the other, and vice versa.

Thus the quantity r_{yx} shows the dependence of one variable upon the other. Sometimes the linear correlation coefficient is defined as the amount of dependence one variable has upon another.

If the variables are not connected by a linear relation, the linear coefficient of correlation should not be used. It has very little meaning when the variables are not connected linearly.

When the observed values of x and y are not very large and the number of observations does not exceed 50 the best formula for computing the correlations coefficient is

$$(10.8) \quad r_{yx} = \frac{\Sigma\bar{x}\bar{y}}{\sqrt{\Sigma\bar{x}^2\Sigma\bar{y}^2}} = \frac{n\Sigma xy - \Sigma x \cdot \Sigma y}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}.$$

If values of x and y are large it will shorten calculations by subtracting provisional means, and then using the new set of data. The value of r, obtained from (10.8), is

$$(10.9) \qquad r_{yx} = \frac{\Sigma\bar{x}\bar{y}}{\sqrt{\Sigma\bar{x}^2 \cdot \Sigma\bar{y}^2}} = \frac{\Sigma\bar{x}'\bar{y}'}{\sqrt{\Sigma\bar{x}'^2 \cdot \Sigma\bar{y}'^2}}$$
$$= \frac{n\Sigma x'y' - (\Sigma x')(\Sigma y')}{\sqrt{[n\Sigma x'^2 - (\Sigma x')^2][n\Sigma y'^2 - (\Sigma y')^2]}},$$

where the primes represent values after provisional means have been subtracted from the variates.

EXAMPLE. Find the coefficient of correlation between weights and right thigh measurements of men if the following measurements are given.

WEIGHT IN POUNDS y	RIGHT THIGH, INCHES x	$y - 140$ y'	$x - 21$ x'	$x'y'$	x'^2	y'^2
139	20.0	− 1	−1.0	+ 1.0		
117	19.0	−23	−2.0	+ 46.0		
150	20.4	+10	−0.6	− 6.0		
166	24.0	+26	+3.0	+ 78.0		
122	19.5	−18	−1.5	+ 27.0		
119	19.5	−21	−1.5	+ 31.5		
146	22.2	+ 6	+1.2	+ 7.2		
137	21.5	− 3	+0.5	− 1.5		
174	24.2	+34	+3.2	+108.8		
141	21.2	+ 1	+0.2	+ 0.2		
		+11	+1.5	+292.2	30 83	3,273

Using (10.9)

$$r_{yx} = \frac{10(292.2) - (1.5)(11)}{\sqrt{[10(30.83) - (1.5)^2]\,[10(3{,}273) - (11)^2]}}$$

$$r_{yx} = 0.9197.$$

OTHER FORMS OF r_{yx}

The correlation coefficient can be written as

$$r_{yx} = \frac{\Sigma xy - nM_xM_y}{n\sigma_x\sigma_y} = \frac{\Sigma x'y' - nM_{x'}M_{y'}}{n\sigma_{x'}\sigma_{y'}}.$$

The standard error of prediction is

$$\sigma_e = \sigma_y\sqrt{1 - r^2};\ \text{hence}\ r^2 = 1 - \frac{\sigma_e^{\,2}}{\sigma_y^{\,2}} = 1 - \frac{\Sigma y^2 - a\Sigma y - b\Sigma xy}{\Sigma \bar{y}^2},$$

which gives r in terms of the summations Σy, Σy^2, $\Sigma\bar{y}^2$, Σxy, and the constants a and b.

$$\sigma_{x-y}^{\,2} = \frac{\Sigma\overline{x - y}^2}{n} = \frac{\Sigma(x - y - M_{x-y})^2}{n} = \frac{\Sigma(x - M_x - y + M_y)^2}{n}$$

$$= \frac{\Sigma(\bar{x} - \bar{y})^2}{n} = \frac{\Sigma\bar{x}^2}{n} - \frac{2\Sigma\bar{x}\bar{y}}{n} + \frac{\Sigma\bar{y}^2}{n} = \sigma_x^{\,2} - \frac{2\Sigma\bar{x}\bar{y}}{n} + \sigma_y^{\,2}.$$

Therefore

$$\frac{2\Sigma\bar{x}\bar{y}}{n} = \sigma_x^{\,2} + \sigma_y^{\,2} - \sigma_{x-y}^{\,2}.$$

Hence

$$r_{yx} = \frac{\sigma_x^{\,2} + \sigma_y^{\,2} - \sigma_{x-y}^{\,2}}{2n\sigma_x\sigma_y}.$$

PROBLEMS

1. Find the linear correlation coefficient between weights and heights, using data on page 151.

2. Do the same for weights and chest measurements, using data on page 157.

3. How do the standard errors of prediction compare with the different correlation coefficients for problems 1 and 2 and that computed on page 152? (Compare pages 157, 161, and 171.)

4. Given the following grades for freshmen in the Iowa Placement Test in mathematics and in their first-semester work in college mathe-

matics. Find the linear correlation coefficient between scores on the test and first-semester grades in mathematics. Choose provisional means.

Iowa Placement Test	First Semester Mathematics Grades	Iowa Placement Test	First Semester Mathematics Grades
40	75	31	75
37	85	50	92
30	72	39	88
47	82	27	75
18	55	41	75
57	88	60	95
28	55	39	60
25	65	42	85
29	88	35	65
39	78	37	74

5. When will $r = o$ represent perfect correlation?

RANK CORRELATION COEFFICIENT

Let x be the rank of the variate instead of the actual value, and let y be the corresponding rank of another variate. If $x = 1$, then the variate corresponding to this x is either the largest or smallest of all the variates. Let us consider rank in descending order. A rank of 2 for a measurement means that the corresponding variate or measurement is second largest. A rank of n means that the corresponding variate is the least in the set of n variates. Here x represents the ranks of one set of variates and y represents the ranks of the corresponding variates in the other set. A man may be the tallest man in a group of men, but his weight may be fourth from the heaviest man in the group. Then $x = 1$ and $y = 4$.

The rank correlation coefficient is the linear correlation coefficient between ranks of the corresponding variates. Let the x's be placed in descending order. The x's will now be the first n integers.

x	y	$D = x - y$
1	3	-1
2	5	-3
3	1	2
4	2	2
.	.	.
.	.	.
.	.	.
n	s	$n - s$

The sum of the x's is equal to the sum of the y's which is also the sum of the first n integers; this is equal to

$$\Sigma x = \Sigma y = \frac{n}{2}(1+n);\text{ also } \Sigma x^2 = \Sigma y^2 = \frac{n(n+1)(2n+1)}{6}$$

$$= 1^2 + 2^2 + \dots + n^2.*$$

The correlation coefficient is, as before,

$$r_{yx} = \frac{\Sigma \bar{x}\bar{y}}{\sqrt{\Sigma \bar{x}^2 \Sigma \bar{y}^2}}.$$

Everything is known except $\Sigma \bar{x}\bar{y} = \Sigma xy - (\Sigma x)(\Sigma y)/n$. From the difference D in the ranks of x and y one can secure a value for $\Sigma \bar{x}\bar{y}$. The first difference is

$$D_1 = x_1 - y_1;$$

therefore

$$D_1^2 = x_1^2 - 2x_1y_1 + y_1^2, \quad \text{and} \quad 2x_1y_1 = x_1^2 - D_1^2 + y_1^2;$$

hence the sum of all the product terms is

$$2\Sigma xy = \Sigma x^2 - \Sigma D^2 + \Sigma y^2 = 2\Sigma x^2 - \Sigma D^2 \text{ and } \Sigma xy = \Sigma x^2 - \Sigma D^2/2.$$

Substitute the values of $\Sigma \bar{x}^2$, $\Sigma \bar{y}^2$, and $\Sigma \bar{x}\bar{y}$ in the formula for r_{yx}; this gives

$$R_{yx} = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

$$= \frac{nn(n+1)(2n+1)/6 - n\Sigma D^2/2 - n^2(n+1)^2/4}{\sqrt{[nn(n+1)(2n+1)/6 - n^2(n+1)^2/4]^2}}$$

$$= 1 - \frac{\Sigma D^2}{n(n+1)[(2n+1)/3 - (n+1)/2]}$$

$$= 1 - \frac{6\Sigma D^2}{n(n+1)(4n+2-3n-3)}.$$

$$(10.10) \quad R_{yx} = 1 - \frac{6\Sigma D^2}{n(n^2-1)}.$$

* This is given in some college algebras.

The symbol R_{yx} will be used to represent the rank correlation coefficient. The rank correlation coefficient can in many instances be calculated much more quickly than the Pearson linear correlation coefficient, for the ΣD^2 can be obtained easily. Differences in the ranks are usually very small and can be squared and summed quickly after the data are arranged in ranks.

Often, R_{yx} gives as much information about the two corresponding sets of variates as is desired.

If there are ties in the ranks, make the ranks so that the sum of the ranks is the same as if there had been no ties. For example, suppose 3 tied for fifth place. The ranks would be 5, 6, and 7, had there been no ties. To make the ranks so that the sum of these ranks is the same as if there had been no ties, make $x_5 = 6$, $x_6 = 6$, $x_7 = 6$. This 6 would be the average of the three ranks if there had been no ties. If x_9 and x_{10} tied for ninth place, make the rank of both 9.5. In this way the sum of the ranks is still $n(n + 1)/2$, but the sum of the squares of the ranks has been slightly changed, but not enough to affect the result much. When there are ties the rank correlation coefficient is not exact or exactly correct, yet it is used as though it were correct. The quantities r_{yx} and R_{yx} are about the same. However, this is not always the case, as the following two series will show:

$$x\text{: } 60,\ 50,\ 40,\ 30,\ 10,$$

$$y\text{: } 100,\ 98,\ 97,\ 3,\ 1.$$

Here the rank correlation coefficient is 1 or perfect, while r_{yx} is not equal to unity.

Example. Let us find the rank correlation coefficient of the following data. The x's have been arranged in descending order.

Freshmen Mathematics Averages	Four-Year Average	Ranks x	Ranks y	D $x - y$	D^2
95	90	1.5	1	0.5	0.25
95	89	1.5	2	−0.5	0.25
92	82	3	6	3.0	9.00
90	83	4	4.5	−0.5	0.25
88	83	5	4.5	0.5	0.25
85	80	6	7.5	−1.5	2.25
82	77	7.5	11	−3.5	12.25
82	79	7.5	9.5	−2.	4.00
78	70	10.5	17.5	−7.	49.00
78	84	10 5	3	7.5	56.25
78	79	10.5	9.5	1	1.00
78	80	10.5	7.5	3	9.00
75	74	13.5	14.5	−1	1.00
75	76	13.5	12.5	1	1.00
72	76	15	12.5	−2.5	6.25
68	74	16.5	14.5	2	4.00
68	70	16.5	17.5	−1	1.00
65	67	18	20	−2	4.00
58	69	19	19	0	0.00
50	72	20	16	−4	16.00
					177.00

$$R_{xy} = 1 - \frac{6\Sigma D^2}{n(n^2 - 1)} = 1 - \frac{6(177)}{20(399)} = 1 - 0.133$$

$$= 0.867,$$

which shows that there is high correlation between what a person does in freshman mathematics and what he will do during his four years in college.

PROBLEMS

1. Find the rank correlation coefficient for the data on page 171 in problem 4. How does the Pearson linear correlation coefficient compare with the rank correlation coefficient?

2. The following figures are the freshmen's mathematics average and their averages made on science courses taken after the freshman year. These data were taken from records at the University of Texas.

Freshmen Mathematics Average	Science Average Above First Year	Freshmen Mathematics Average	Science Average Above First Year
68	73	68	72
75	65	85	78
78	77	72	85
95	95	88	89
82	75	85	81
65	74	85	82
78	87	90	92
85	90	79	76
75	67	78	75
72	81	82	87
82	82	92	93
72	73	78	75
72	69	76	79

Find the rank correlation coefficient for these data. Does ability in mathematics help one in sciences?

GRAPHIC MEANING OF THE CORRELATION COEFFICIENT

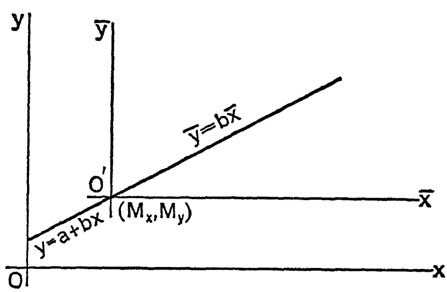


Fig. 10.1.—Predicting line with respect to the x and y axes and with respect to the $\bar{x}$ and $\bar{y}$ axes.

The above graph, Fig. 10.1, shows the predicting equation plotted with respect to the x and y axes and with respect to the $\bar{x}$ and $\bar{y}$ axes. The predicting equation, with the x and y axes as reference lines, is $y = a + bx$ and with the $\bar{x}$ and $\bar{y}$ axes, after the origin has been changed to (M_x, M_y), is $\bar{y} = b\bar{x}$, or $\bar{y} = \frac{\sigma_y}{\sigma_x} \cdot r_{yx} \cdot \bar{x}$.

Let $$t_x = \frac{\bar{x}}{\sigma_x} \quad \text{and} \quad t_y = \frac{\bar{y}}{\sigma_y},$$

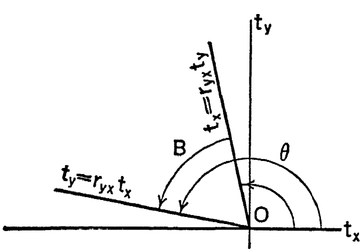
then the equation of this line becomes

(10.11) $$t_y = r_{yx} \cdot t_x,$$

which gives the line of prediction in terms of standard units and the correlation coefficient. These units are abstract and do not depend upon the units of x and y given originally. In this equation the correlation coefficient r_{yx} is the slope of the predicting line or what is sometimes called the line of regression. If $0 \leq r \leq 1$, the tangent of the angle which this line makes with the horizontal axis is always less than or equal to 1. Call this angle θ. In this case θ will be less than 45°. If r is negative, θ will satisfy the inequalities, $135° \leqq \theta \leqq 180$. The angles θ are shown in Figs. 10.2 and 10.3.

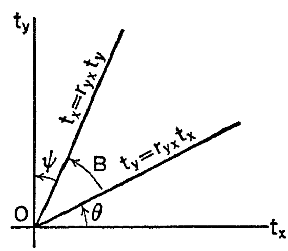
Figures 10.2 and 10.3 show parts of the predicting line plotted in standard units and the values of θ when r is positive and negative.

FIG. 10.2



$0 \geqq r_{yx} \geqq -1.$

FIG 10.3



$0 \leqq r_{yx} \leqq 1.$

If x is predicted from y, another predicting equation is obtained. Let it be

$$x = c + dy, \quad \text{or} \quad \bar{x} = d\bar{y}, \quad \text{or} \quad t_x = r_{xy}t_y$$

Let B be the angle between the predicting line for y and the predicting line for x. Examine the above figures. If B is zero there is perfect correlation. The size of B gives information concerning the correlation between the two variables. If r is positive, $0 \leqq B \leqq 90$. If B is large there is very little correlation between the two variables.

PROBLEM

1. Find the predicting equation for predicting heights from weights from data on page 151. Plot the predicting equation for weights from heights and the predicting equation for heights from weights. Plot these two equations on the same graph in terms of the deviations from the means. Plot these two lines on another graph in terms of standard units. Find the size of θ, of Ψ, and of angle B.

Correlation Table of Shoulder and

Shoulder

Weight Measurements (y) Pounds

Classes	Class Mark	f	d	$d\,f$	$d^2 \cdot f$	13 95 to 14 45	14 45 to 14 95	14.95 to 15 45	15 45 to 15 95
						14 20	14.70	15 20	15 70
						6	15	49	68
						− 5	− 4	− 3	− 2
						− 30	− 60	−147	−136
						150	240	441	272
209 5 to 219 5	214 5	1	7	7	49				
199 5 to 209 5	204 5	2	6	12	72				
189 5 to 199.5	194 5	7	5	35	175				
179 5 to 189 5	184 5	10	4	40	160				
169 5 to 179 5	174 5	24	3	72	216				
159 5 to 169.5	164 5	36	2	72	144		−8 **1** −8		−4 **1** −4
149 5 to 159 5	154.5	97	1	97	97			−3 **1** −3	−2 **2** −4
139 5 to 149.5	144.5	157	0	0				0 **5** 0	0 **9** 0
129 5 to 139 5	134.5	174	−1	−174	174	5 **1** 5	4 **2** 8	3 **11** 33	2 **14** 28
119 5 to 129.5	124.5	129	−2	−258	516		8 **5** 40	6 **17** 102	4 **23** 92
109.5 to 119.5	114.5	67	−3	−201	603	15 **1** 15	12 **4** 48	9 **11** 99	6 **17** 102
99.5 to 109.5	104.5	13	−4	− 52	208	20 **3** 60	16 **3** 48	12 **4** 48	8 **2** 16
89.5 to 99.5	94.5	1	−5	− 5	25	25 **1** 25			
Total		718		−355 $\Sigma y'$	2,439 $\Sigma y'^2$	105	136	279	230

WEIGHT MEASUREMENTS OF UNIVERSITY MEN

MEASUREMENTS (x) INCHES

15 95 to 16.45	16 45 to 16 95	16 95 to 17 45	17 45 to 17 95	17 95 to 18 45	18 45 to 18 95	18 95 to 19 45	19 45 to 19 95	Totals
16 20	16 70	17 20	17 70	18 20	18 70	19 20	19 70	
170	129	141	74	39	14	12	1	
− 1	0	1	2	3	4	5	6	718
−170	0	141	148	117	56	60	6	−15 $\Sigma x'$
170		141	296	351	224	300	36	2,621 $\Sigma x'^2$
						35 **1** 35		35
					24 **1** 24	30 **1** 30		54
	0 **1** 0	5 **1** 5	10 **2** 20	15 **1** 15	20 **1** 20	25 **0** 0	30 **1** 30	90
	0 **1** 0	4 **1** 4	8 **1** 8	12 **2** 24	16 **3** 48	20 **2** 40		124
−3 **2** −6	0 **2** 0	3 **3** 9	6 **4** 24	9 **6** 54	12 **3** 36	15 **4** 60		177
−2 **2** −4	0 **4** 0	2 **11** 22	4 **9** 36	6 **6** 36	8 **1** 8	10 **1** 10		96
−1 **12** −12	0 **8** 0	1 **20** 20	2 **30** 60	3 **17** 51	4 **4** 16	5 **3** 15		143
0 **19** 0	0 **48** 0	0 **47** 0	0 **21** 0	0 **7** 0	0 **1** 0	0 0	0	
1 **61** 61	0 **38** 0	−1 **40** −40	−2 **7** −14	**0**				81
2 **47** 94	0 **22** 0	−2 **15** −30						298
3 **27** 81	0 **4** 0	−3 **3** −9						336
	0 **1** 0							172
								25
214		−19	134	180	152	190	30	**1,631** $\Sigma x'y'$

A METHOD FOR COMPUTING THE CORRELATION COEFFICIENT WHEN THERE IS A LARGE NUMBER OF ITEMS

When there is a large number of variates for the variables under consideration it would require a great deal of trouble to find the sum of the product terms by the method given above. One way of reducing the amount of computing is to group the variates for the two sets of data. The classes for the x variates may not be the same size as the classes for the y variates. After the data have been grouped into their respective classes the class marks are then the representatives of the variates falling in the different classes.

One of the best ways of understanding this method is to go through an example. On page 178 are a two-way correlation table and the necessary calculations for finding the coefficient of correlation. The quantity x represents shoulder measurements of university men; y represents their weights. Shoulder measurements were put into classes of $\frac{1}{2}$ inch; the weights were grouped into classes of 10 pounds. The provisional mean for the x's is 16.70; that for the y's is 144.5 pounds. These are indicated by the heavy lines near the middle of the table. The figures in the center of the squares represent frequencies of the double classes. For example, there are 9 men with shoulder measurements between 17.45 and 17.95 inches and with weights between 159.5 and 169.5 pounds; there are 23 men with shoulder measurements between 15.45 and 15.95 inches and with weights between 119.5 and 129.5 pounds, etc. There are 3 men in the double class 18.45–18.95 inches and 169.5–179.5 pounds. This class is 4 units from the provisional mean for the x's and 3 units from the provisional mean for the y's. The product of these is 12 double class units or product units. This 12 appears in the upper right-hand corner of this square. Since the frequency of this double class is 3, there are hence 3×12 double units for this class. This 36 is placed in the lower left-hand corner of the square. The double class 15.95–16.45 inches and 169.5–179.5 pounds is -1 class unit from the provisional mean for the x's and $+3$ class units from the provisional mean for the y's, hence this class is -3 double class units from the origin (16.70 inches, 144.5 pounds). Since the frequency of this class is 2, there are -6 double units for this class. This -6 appears in the lower left-hand corner of this square. The sum of all the numbers in the lower left-hand corners of the squares gives the cross product

term $\Sigma x'y'$. In this case $\Sigma x'y' = 1{,}631$, which is found in the extreme lower right-hand square. It is the sum of the product terms for the squares in the table, and can be found by adding the product terms vertically or horizontally. The entire object of the above correlation table is to find the cross product term $\Sigma x'y'$.

The column and the row $d \cdot f$ enable one to secure the means for the two variables and also the sums $\Sigma x'$ and $\Sigma y'$. The column and row d^2f enable one to secure the second moments for the two sets and also the sums $\Sigma x'^2$ and $\Sigma y'^2$, from which the standard deviations can be obtained.

From the correlation table one gets

$$\Sigma y' = -355,\ \Sigma x' = -15,\ \Sigma y'^2 = 2{,}439,\ \Sigma x'^2 = 2{,}621,$$

$$\Sigma x'y' = 1{,}631,$$

from which the following are obtained

$$\Sigma \bar{y}'^2 = 2{,}263.48,\ \Sigma \bar{x}'^2 = 2{,}620.69,\ \Sigma \bar{x}'\bar{y}' = 1{,}623.58.$$

Since the data are grouped, corrections must be made in finding the standard deviations.

$$\sigma_y = \left(\sqrt{\frac{2{,}263.48}{718} - 0.08333}\right)10 = 17.52,$$

$$\sigma_x = \left(\sqrt{\frac{2{,}620.69}{718} - 0.08333}\right)0.5 = 0.94.$$

Therefore the correlation coefficient is

$$r_{yx} = \frac{\Sigma \bar{x}\bar{y}}{n\sigma_x \cdot \sigma_y} = \frac{[1{,}623.58]10(0.5)}{718(17.52)(0.94)} = \frac{8{,}117.90}{11{,}824.45} = 0.686.$$

PROBLEMS

1. Given the following data for freshmen English average and the average made in all subjects during the 4 years in college, for women at the University of Texas. The grades were made to the nearest unit. Group the data in 5's and find the correlation coefficient between freshmen English average and 4-year average in all subjects. Use 64.5–69.5 as the first class for English grades and 59.5–64.5 for the first class for the four-year averages.

English Average x	4-Year Average y	English Average x	4-Year Average y	English Average x	4-Year Average y	English Average x	4-Year Average y	English Average x	4-Year Average y
85	84	85	81	92	86	75	81	68	76
78	74	85	84	68	74	78	70	82	91
75	79	82	85	88	81	85	77	82	76
68	73	75	72	85	70	80	78	78	73
82	71	92	94	85	78	75	75	72	77
82	82	88	89	71	71	72	73	85	79
85	84	92	92	65	68	79	75	85	86
90	72	82	74	88	82	75	75	78	76
92	92	68	80	78	80	75	78	85	79
88	82	65	67	78	80	88	85	95	87
72	78	65	69	95	88	68	83	68	76
85	80	82	80	95	85	65	77	72	80
78	71	85	85	88	90	82	85	72	80
81	85	82	80	65	74	75	79	75	79
68	62	85	87	75	76	68	71	78	85
72	76	65	71	92	84	82	83	65	80
68	83	85	87	75	82	75	81	95	90
85	85	82	74	95	88	88	84	85	85
82	76	78	83	68	72	85	90		
75	78	72	83	75	71	88	86		

2. Find the correlation coefficient between weights and right thigh measurements for the data given in the following table. Assume a linear relation. Find the predicting equation for predicting weights from right thigh measurements. Find the mean weight for each group of thigh measurements. Plot these means on the same plot with the predicting line. Draw lines parallel to the predicting line at a distance of 1 σ_e from it. What percentage of these means falls in this band? What value does the predicting line give for any thigh measurement?

CORRELATION TABLE

Weights and Right Thigh Measurements of Freshmen at University of Michigan

Weights to Nearest Pound

Right thigh measurements to nearest 1/10 inch

	79.5 89.5	99.5 109 5	109.5 119.5	119 5 129.5	129.5 139.5	139.5 149.5	149 5 159.5	159.5 169.5	169.5 179 5	179 5 189.5	189.5 199.5	199.5 209 5	209 5 219.5	Total
15.95–	......	5	3	......	......	......	1	......	......	......	......	......		9
16.95–	1	7	21	15	3	......	......	......	......	......	......	......	......	47
17.95–	......	6	37	60	42	6	2	......	......	......	......	......	......	153
18.95–	......	1	18	77	97	56	12	......	......	......	......	......	......	261
19.95–	......	......	1	23	96	78	46	9	2	......	......	......	......	255
20.95–	......	......	......	4	19	53	42	21	6	......	......	......	......	145
21.95–	......	......	......	......	3	15	28	23	5	2	......	......	......	76
22.95–	......	......	......	......	......	2	7	8	11	3	1	1	......	33
23.95–	......	......	......	......	......	......	2	......	3	2	......	......	1	8
24.95–	......	......	......	......	......	......	......	1	1	3	3	3	1	12
25.95–	......	......	......	......	......	......	......	......	......	1	......	......	......	1
Total	1	19	80	179	260	210	140	62	28	11	4	4	2	1,000

THE VALUE OF THE CORRELATION COEFFICIENT IN PREDICTING

Consider the following table for various values of the correlation coefficient:

r_{yx}	$1 - r^2$	$\sqrt{1 - r^2}$	$1 - \sqrt{1 - r^2}$
0.1	0.99	0.995	0 005
.2	.96	.980	.020
.3	.91	.954	.046
.4	.84	.947	.053
.5	.75	.886	.134
.6	.64	.800	.200
.7	.51	.714	.286
.8	.36	.600	.400
.9	.19	.436	.564
.92	.15	.392	.608
.94	.12	.341	.659
.96	.08	.280	.820
.98	.04	.198	.811
1.00	0 00	0.000	1.000

The question arises as to the importance of r_{yx} in predicting values of one variable when values of the others are known. If the correlation coefficient $r_{yx} = 0$, the predicting equation reduces to $y = M_y$, which means that for each value of x the predicted value of y is always equal to M_y. This will be considered a mere guess for the value of y. In this case the standard error of prediction is $\sigma_e = \sigma_y$. If $r_{yx} = 0.5$, then $\sigma_e = 0.866\sigma_y$; this value of σ_e is 0.134 of what it was when y was predicted by a mere guess. If all the y's had been multiplied by 0.866, then $\sigma_e = 0.866\ \sigma_y$ and the mean of these y's would have been $0.866M_y$. From the information available the best predicting equation would be $y = 0.866M_y$, which is absurd, since y and x are connected by some equation and the value of y cannot be constant for all values of x. This shows that, if one knows that two variables are connected in such a way as to give a linear correlation coefficient of 0.5 and nothing more is known about the two variables, the best predicting value of y would be $0.866M_y$, where M_y is the mean of all values of y. This goes to show that the correlation coefficient is not very helpful in predicting values of one variable from values of the other.* The essentials

* The value of r can be used in predicting if the variables are in standard units.

in predicting are the predicting equation, $y = a + bx$, and the standard error of prediction.

The quantity $1 - \sqrt{1 - r^2}$ shows how much worse off one is by using values of x and the value of the correlation for predicting values of y.

If $r = 0.6$, how much has the standard error of prediction been reduced from that obtained by making a mere guess, that is, when $y = My$ is used as the predicting equation? What is $1 - \sqrt{1-r^2}$?

COEFFICIENT OF DETERMINATION

It is interesting to note that the equation $\sigma_e = \sigma_y\sqrt{1 - r_{yx}^2}$ may be solved for r, giving

$$r_{yx} = \sqrt{1 - \frac{\sigma_e^2}{\sigma_y^2}}. \tag{10.12}$$

This shows that as σ_e becomes smaller the correlation coefficient r_{yx} becomes larger, and as σ_e approaches σ_y the value of r approaches 0.

The value of r^2 is sometimes called the coefficient of determination for it measures the per cent of the variability of the dependent set of variates determined from the independent. This coefficient of determination is

$$r_{yx}^2 = \frac{\sigma_y^2 - \sigma_e^2}{\sigma_y^2}.$$

It is really the ratio of the difference between σ_y^2 and σ_e^2 to σ_y^2. If σ_e^2 represents the variability of the values of the variates of the dependent variable about the predicting equation and σ_y^2 represents the variability of these variates about their mean, then $\sigma_y^2 - \sigma_e^2$ represents the part of the variability of the dependent variable accounted for by the variability in the independent variable. This coefficient, r^2, is the ratio of the variability of y due to the variability in x and the variability of y, since $\sigma_y^2 = \sigma_y^2 (1 - r^2) + r^2 \sigma_y^2$.

PROBLEMS

1. Given that y and x are connected linearly, and the following:

$$M_x = 34.7, \quad M_y = 67.3, \quad \sigma_x = 1.9, \quad \sigma_y = 2.3, \quad r_{yx} = 0.01.$$

Write the predicting equation; discuss it as a predicting equation in comparison with $y = M_y$ as a predicting equation. Has the correlation coefficient helped in predicting values of y? Examine b.

2. Suppose that waist measurements, x, and weights, y, are connected linearly and that $r_{yx} = 0.87$, $\sigma_y = 19.3$ pounds, $\sigma_x = 1.4$ inches. Find the value of $\bar{y}$ when x is 1.3 standard deviations above the mean of the x's.

3. Using the data in problem 2, find the predicted value of y if in addition it is known that $M_y = 139.6$ pounds, $M_x = 29.8$ inches, and x is equal to 30.9 inches.

4. If $\sigma_e = 15$ pounds, and $r = 0.8$, find σ_y.

MULTIPLE CORRELATION

Multiple correlation shows in a measure how one variable depends upon two or more variables. The multiple correlation coefficient measures the amount of dependence one variable has upon two or more variables when the variables are connected in pairs by linear relations and a "plane" is the surface which fits the data best. If weights, heights, and chest measurements are connected linearly, the multiple correlation coefficient gives, to some extent, the amount of dependence of the weight of an individual upon his height and chest measurement. This coefficient will be derived by a method similar to that for obtaining the Pearson linear correlation coefficient.

Let y, x, and z be connected by the linear relation

$$y = a + bx + cz, \quad \text{or} \quad \bar{y} = b\bar{x} + c\bar{z}.$$

The normal equations are

$$\Sigma\bar{x}^2 \cdot b + \Sigma\bar{x}\bar{z} \cdot c = \Sigma\bar{x}\bar{y}$$

$$\Sigma\bar{x}\bar{z}b + \Sigma\bar{z}^2 \cdot c = \Sigma\bar{z}\bar{y},$$

from which

$$b = \frac{\Sigma\bar{z}^2\Sigma\bar{x}\bar{y} - \Sigma\bar{x}\bar{z}\Sigma\bar{z}\bar{y}}{\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{x}\bar{z})^2}, \quad \text{and} \quad c = \frac{\Sigma\bar{x}^2\Sigma\bar{z}\bar{y} - \Sigma\bar{z}\bar{x}\Sigma\bar{x}\bar{y}}{\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{z}\bar{x})^2}.$$

The linear relation is

$$(10.13) \quad \bar{y} = \left[\frac{\Sigma\bar{z}^2\Sigma\bar{x}\bar{y} - \Sigma\bar{z}\bar{x}\Sigma\bar{z}\bar{y}}{\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{x}\bar{z})^2}\right]\bar{x} + \left[\frac{\Sigma\bar{x}^2\Sigma\bar{z}\bar{y} - \Sigma\bar{z}\bar{x}\Sigma\bar{x}\bar{y}}{\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{x}\bar{z})^2}\right]\bar{z}.$$

The standard error of prediction is as before

$$\sigma_e = \sqrt{[\Sigma\bar{y}^2 - b\Sigma\bar{x}\bar{y} - c\Sigma\bar{z}\bar{y}]/n},$$

or

$$\sigma_e = \sqrt{\frac{\left\{\Sigma\bar{y}^2 - \dfrac{[\Sigma\bar{z}^2\Sigma\bar{x}\bar{y} - \Sigma\bar{x}\bar{z}\Sigma\bar{z}\bar{y}]\Sigma\bar{x}\bar{y}}{\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{x}\bar{z})^2} - \dfrac{[\Sigma\bar{x}^2\Sigma\bar{z}\bar{y} - \Sigma\bar{x}\bar{z}\Sigma\bar{x}\bar{y}]\,\Sigma\bar{z}\bar{y}}{\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{x}\bar{z})^2}\right\}}{n}}$$

$$= \sqrt{\left[\Sigma\bar{y}^2 - \frac{\Sigma\bar{z}^2(\Sigma\bar{x}\bar{y})^2 - \Sigma\bar{x}\bar{z}\Sigma\bar{z}\bar{y}\Sigma\bar{x}\bar{y} + \Sigma\bar{x}^2(\Sigma\bar{z}\bar{y})^2 - \Sigma\bar{x}\bar{z}\Sigma\bar{x}\bar{y}\Sigma\bar{z}\bar{y}}{\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{x}\bar{z})^2}\right]\frac{1}{n}}$$

$$= \sqrt{\frac{\Sigma\bar{y}^2}{n}\left[1 - \frac{\Sigma\bar{z}^2(\Sigma\bar{x}\bar{y})^2 + \Sigma\bar{x}^2(\Sigma\bar{z}\bar{y})^2 - 2\Sigma\bar{x}\bar{z}\Sigma\bar{x}\bar{y}\Sigma\bar{z}\bar{y}}{[\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{x}\bar{z})^2]\Sigma\bar{y}^2}\right]}$$

$$= \sigma_y \sqrt{1 - R_{y.xz}{}^2},$$

where

$$(10.14) \qquad R_{y.xz} = \sqrt{\frac{\Sigma\bar{z}^2(\Sigma\bar{x}\bar{y})^2 - 2\Sigma\bar{x}\bar{y}\Sigma\bar{x}\bar{z}\Sigma\bar{z}\bar{y} + \Sigma\bar{x}^2(\Sigma\bar{z}\bar{y})^2}{[\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{x}\bar{z})^2]\Sigma\bar{y}^2}},$$

which is called the multiple correlation coefficient.

The standard error of estimate may be written as

$$\sigma_e = \sigma_y \sqrt{1 - \frac{\dfrac{(\Sigma\bar{x}\bar{y})^2}{\Sigma\bar{x}^2\Sigma\bar{y}^2} + \dfrac{(\Sigma\bar{z}\bar{y})^2}{\Sigma\bar{z}^2\Sigma\bar{y}^2} - \dfrac{2\Sigma\bar{x}\bar{z}\Sigma\bar{z}\bar{y}\Sigma\bar{x}\bar{y}}{\Sigma\bar{z}^2\Sigma\bar{y}^2\Sigma\bar{x}^2}}{1 - \dfrac{(\Sigma\bar{x}\bar{z})^2}{\Sigma\bar{x}^2\Sigma\bar{z}^2}}}$$

or

$$(10.15) \qquad \sigma_e = \sigma_y \sqrt{1 - \frac{r_{yx}{}^2 - 2r_{yx}r_{xz}r_{zy} + r_{yz}{}^2}{1 - r_{xz}{}^2}};$$

hence the multiple correlation coefficient is equal to

$$(10.16) \qquad R_{y.xz} = \sqrt{\frac{r_{yx}{}^2 - 2r_{yx}r_{yz}r_{xz} + r_{zy}{}^2}{1 - r_{xz}{}^2}}.$$

The multiple correlation coefficient is given in (10.14) in terms of the fundamental summations; it is given in terms of the Pearson linear correlation coefficients in (10.16). Formula (10.14) is the most practical one to use.

As in the Pearson linear correlation coefficient, the multiple correlation coefficient cannot be greater than 1 or less than −1. It is always considered to be positive.

The symbol $R_{y.xz}$ means the multiple correlation coefficient between y and x and z. It arose from a linear relation connecting y with x and z. The points are scattered about a plane in three dimensions.

EXAMPLE. The multiple correlation coefficient between weight, and heights, and right thigh measurements, by using data on page 165, and formula (10.14), is

$$R_{wt\ ht\ ch} = \sqrt{\frac{(30.605)\ (265.5)^2 - 2(265.5)\ (20.3)\ (290\ 6) + 66\ 98(290.6)^2}{[(66.98)\ (30.605) - (20.3)^2]\ (3{,}260.9)}}$$

$$= 0.947.$$

PROBLEMS

1. If the multiple correlation coefficient is 1, and $r_{yx} = 0.1$, $r_{zy} = 0.2$, find r_{xz}, and discuss your results.

2. (*a*) If $r_{xy} = r_{zy} = 0$, what does $R_{y \cdot xz}$ equal? Discuss this situation. (*b*) If $r_{xy} = R_{y\ xz} = 0$, what does r_{zy} equal? Discuss this situation.

3. Discuss the situation when $r_{xy} = r_{zy} = 1$ in formula (10.15). Make a drawing to illustrate.

4. Find the multiple correlation coefficients between weights, heights, and chest measurements, if it is assumed that $y = a + bx + cz$, where y represents weights, x heights, and z chest measurements. Use the data on pages 151 and 157.

5. Assume that the four-year average grade is linearly connected with freshman mathematics grades and freshman English grades. Find the multiple correlation coefficient between four-year average and freshman mathematics and English grades if the following data are given for men. These data were taken from records at the University of Texas.

Mathematics Averages	English Averages	4-Year Averages	Mathematics Averages	English Averages	4-Year Averages
68	75	74	79	68	72
75	82	74	68	72	68
78	82	84	85	75	79
95	88	90	72	82	78
82	72	79	88	68	77
65	68	67	85	86	84
78	92	80	85	85	85
75	68	76	90	88	90
85	81	80	79	82	79
72	75	76	78	68	70
82	82	78	50	65	72
72	60	72	82	75	85
82	78	80			

6. Discuss the situation when $r_{xz} = 1$, in formula (10.16).
7. When will $r_{xy} = r_{xz} = r_{zy} = R_{y\,xz}$?
8. Given that $r_{yx} = 0.62$, $r_{yz} = 0.94$, $r_{xz} = 0.34$; find $R_{y\,xz}$.
9. Express the predicting equation $\bar{y} = b\bar{x} + c\bar{z}$ in terms of standard deviations and correlation coefficients.
10. Find $R_{y.xz}$ when $r_{xy} = r_{xz} = r_{zy} = 0.999$.

MULTIPLE CORRELATION WHEN THERE ARE MORE THAN THREE VARIABLES

Assume that the variables $y, x_1, x_2, \ldots, x_s$ are connected by the linear relation

$$\bar{y} = c_1\bar{x}_1 + c_2\bar{x}_2 + \ldots + c_s\bar{x}_s.$$

The normal equations are

$$\begin{array}{l}
c_1\Sigma\bar{x}_1^2 + c_2\Sigma\bar{x}_1\bar{x}_2 + \ldots + c_s\Sigma\bar{x}_1\bar{x}_s = \Sigma\bar{x}_1\bar{y}, \\
c_1\Sigma\bar{x}_1\bar{x}_2 + c_2\Sigma\bar{x}_2^2 + \ldots + c_s\Sigma\bar{x}_2\bar{x}_s = \Sigma\bar{x}_2\bar{y}, \\
c_1\Sigma\bar{x}_1\bar{x}_3 + c_2\Sigma\bar{x}_2\bar{x}_3 + \ldots + c_s\Sigma\bar{x}_3\bar{x}_s = \Sigma x_3\bar{y}, \\
\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
c_1\Sigma\bar{x}_1\bar{x}_s + c_2\Sigma\bar{x}_2\bar{x}_s + \ldots + c_s\Sigma\bar{x}_s^2 = \Sigma\bar{x}_s\bar{y},
\end{array}$$

from which

$$c_1 = \frac{\begin{vmatrix}
\Sigma\bar{x}_1\bar{y} & \Sigma\bar{x}_1\bar{x}_2 & \Sigma\bar{x}_1\bar{x}_3 & \ldots\ldots & \Sigma\bar{x}_1\bar{x}_s \\
\Sigma\bar{x}_2\bar{y} & \Sigma\bar{x}_2^2 & \Sigma\bar{x}_2\bar{x}_3 & \ldots\ldots & \Sigma\bar{x}_2\bar{x}_s \\
\Sigma\bar{x}_3\bar{y} & \Sigma\bar{x}_2\bar{x}_3 & \Sigma\bar{x}_3^2 & \ldots\ldots & \Sigma\bar{x}_3\bar{x}_s \\
\ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\
\Sigma\bar{x}_s\bar{y} & \Sigma\bar{x}_2\bar{x}_s & \Sigma\bar{x}_3\bar{x}_s & & \Sigma\bar{x}_s^2
\end{vmatrix}}{\begin{vmatrix}
\Sigma\bar{x}_1\bar{x}_1 & \Sigma\bar{x}_1\bar{x}_2 & \Sigma\bar{x}_1\bar{x}_3 & \ldots\ldots & \Sigma\bar{x}_1\bar{x}_s \\
\Sigma\bar{x}_1\bar{x}_2 & \Sigma\bar{x}_2\bar{x}_2 & \Sigma\bar{x}_2\bar{x}_3 & \ldots\ldots & \Sigma\bar{x}_2\bar{x}_s \\
\Sigma\bar{x}_1\bar{x}_3 & \Sigma\bar{x}_2\bar{x}_3 & \Sigma\bar{x}_3\bar{x}_3 & \ldots\ldots & \Sigma\bar{x}_3\bar{x}_s \\
\ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\
\Sigma\bar{x}_1\bar{x}_s & \Sigma\bar{x}_2\bar{x}_s & \Sigma\bar{x}_3\bar{x}_s & \ldots\ldots & \Sigma\bar{x}_s\bar{x}_s
\end{vmatrix}}$$

Remembering that $\Sigma\bar{x}_i\bar{x}_j = n\sigma_{x_i}\cdot\sigma_{x_j}\cdot r_{x_ix_j}$ the quantity c_1 becomes

$$c_1 = \frac{\begin{vmatrix} n\sigma_{x_1}\sigma_y r_{x_1y} & n\sigma_{x_1}\sigma_{x_2} r_{x_1x_2} & \cdots\cdots & n\sigma_{x_1}\sigma_{x_s} r_{x_1x_s} \\ n\sigma_{x_2}\sigma_y r_{x_2y} & n\sigma_{x_2}^2\, r_{x_2x_2} & \cdots\cdots & n\sigma_{x_2}\sigma_{x_s} r_{x_2x_s} \\ n\sigma_{x_3}\sigma_y r_{x_3y} & n\sigma_{x_2}\sigma_{x_3} r_{x_2x_3} & \cdots\cdots & n\sigma_{x_3}\sigma_{x_s} r_{x_3x_s} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ n\sigma_{x_s}\sigma_y r_{x_sy} & n\sigma_{x_2}\sigma_{x_s} r_{x_2x_s} & \cdots\cdots & n\sigma_{x_s}^2\, r_{x_sx_s} \end{vmatrix}}{\begin{vmatrix} n\sigma_{x_1}^2\cdot 1 & n\sigma_{x_1}\sigma_{x_2} r_{x_1x_2} & \cdots\cdots & n\sigma_{x_1}\sigma_{x_s} r_{x_1x_s} \\ n\sigma_{x_1}\sigma_{x_2} r_{x_1x_2} & n\sigma_{x_2}^2\cdot 1 & \cdots\cdots & n\sigma_{x_2}\sigma_{x_s} r_{x_2x_s} \\ n\sigma_{x_1}\sigma_{x_3} r_{x_2x_3} & n\sigma_{x_2}\sigma_{x_3}\cdot r_{x_2x_3} & \cdots\cdots & n\sigma_{x_3}\sigma_{x_s} r_{x_3x_s} \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ n\sigma_{x_1}\sigma_{x_s} r_{x_1x_s} & n\sigma_{x_2}\sigma_{x_s} r_{x_2x_s} & \cdots\cdots & n\sigma_{x_s}^2\cdot 1 \end{vmatrix}}$$

$$= \frac{\sigma_y \begin{vmatrix} r_{x_1y} & r_{x_1x_2} & r_{x_1x_3} & \cdots & r_{x_1x_s} \\ r_{x_2y} & r_{x_2x_2} & r_{x_2x_3} & \cdots & r_{x_2x_s} \\ r_{x_3y} & r_{x_3x_2} & r_{x_3x_3} & \cdots & r_{x_3x_s} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_sy} & r_{x_2x_s} & r_{x_3x_s} & \cdots & r_{x_sx_s} \end{vmatrix}}{\sigma_{x_1} \begin{vmatrix} 1 & r_{x_1x_2} & r_{x_1x_3} & \cdots & r_{x_1x_s} \\ r_{x_1x_2} & 1 & r_{x_2x_3} & \cdots & r_{x_2x_s} \\ r_{x_2x_3} & r_{x_2x_3} & 1 & \cdots & r_{x_3x_s} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{x_1x_s} & r_{x_2x_s} & r_{x_3x_s} & \cdots & 1 \end{vmatrix}}$$

$$C_1 = \frac{\sigma_y\cdot D_1}{\sigma_{x_1}\cdot D_0},$$

where D_1 is equal to the determinant in the numerator and D_0 is the determinant in the denominator.

In a similar way it can be shown that

$$c_i = \frac{\sigma_y\cdot D_i}{\sigma_{x_i}\cdot D_0}.$$

Hence the predicting equation becomes

$$\bar{y} = \sum \frac{\sigma_y}{\sigma_{x_i}}\cdot\frac{D_i}{D_0}\cdot\bar{x}_i, \tag{10.17}$$

or

$$y = M_y - \frac{\sigma_y}{D_0}\cdot\sum\frac{D_iM_{x_i}}{\sigma_{x_i}} + \frac{\sigma_y}{D_0}\cdot\sum\frac{D_i}{\sigma_{x_i}}x_i.$$

The standard error of estimate is

$$(10.18)\qquad \sigma_e = \sigma_y \sqrt{1 - \frac{1}{n\sigma_y D_0} \sum_{i=1}^{s} \frac{D_i}{\sigma_{x_i}} \sum_{k=1}^{n} \bar{x}_j \bar{y},} \qquad j = i.$$

$$(10.19)\qquad R_{y\,x_1x_2\cdots x_s} = \sqrt{\frac{1}{D_0} \sum_{i=1}^{s} D_i \cdot r_{x_i y},}$$

which is the multiple correlation coefficient between y and the other variables.

EXAMPLE. Given that $\bar{y} = c_1\bar{x}_1 + c_2\bar{x}_2 + c_3\bar{x}_3$, and

$$M_y = 36.6, \quad \sigma_y = 1.9, \quad r_{yx_1} = .52, \quad r_{x_1x_3} = 0.43,$$

$$M_{x_1} = 13.2, \quad \sigma_{x_1} = 0.8, \quad r_{yx_2} = .74, \quad r_{x_2x_3} = 0.51.$$

$$M_{x_2} = 20.2, \quad \sigma_{x_2} = 1.2, \quad r_{xy_3} = 0.80,$$

$$M_{x_3} = 29.4, \quad \sigma_{x_3} = 1.7, \quad r_{x_1x_2} = 0.93;$$

Find the predicting equation and the standard error of prediction.

$$c_1 = \frac{(1.9)\begin{vmatrix} 0.52 & 0.93 & 0.43 \\ 0.74 & 1 & 0.51 \\ 0.80 & 0.51 & 1 \end{vmatrix}}{(0.8)\begin{vmatrix} 1 & 0.93 & 0.43 \\ 0.93 & 1 & 0.51 \\ 0.43 & 0.51 & 1 \end{vmatrix}} = \frac{(1.9)\,(-0.11)}{(0.8)\,(0.10)} = -2.63,$$

$$c_2 = \frac{(1.9)}{1.2} \cdot \frac{\begin{vmatrix} 1 & 0.52 & 0.43 \\ 0.93 & 0.74 & 0.51 \\ 0.43 & 0.80 & 1 \end{vmatrix}}{0.10} = \frac{(1.9)\,(+0.15)}{(1.2)\,(0.10)} = +2.38,$$

$$c_3 = \frac{(1.9)}{1.7} \cdot \frac{\begin{vmatrix} 1 & 0.93 & 0.52 \\ 0.93 & 1 & 0.74 \\ 0.43 & 0.51 & 0.80 \end{vmatrix}}{0.10} = \frac{(1.9)\,(+0.05)}{(1.7)\,(0.10)} = 0.56.$$

The predicting equation is

$$\bar{y} = -2.63\,\bar{x}_1 + 2.38\,\bar{x}_2 + 0.56\,\bar{x}_3.$$

$$\sigma_e = (1.9)\sqrt{1 - \frac{1}{0\ 10}\,[(-0.11)\ (0.52) + (0.15)\ (0.74) + (0.05)\ (0.80)]}$$

$$= 0.48.$$

The multiple correlation coefficient is

$$R_{y\,x_1x_2x_3} = \sqrt{\frac{1}{0\ 1}\,[(-0.11)\ (0.52) + (0.15)\ (0.74) + (0.05)\ (0.80)]}$$

$$= 0.97.$$

PROBLEMS

1. Using the data on pages 151, 157, and 161 concerning weights, heights, chest measurements, and right thigh measurements, find the multiple correlation, $R_{\text{wt. ht ch th.}}$ by means of the correlation coefficients. Write D_0 in terms of the r's and find its numerical value.

2. Derive a formula for c_5 as was done for c_1.

3. Assume that weight is connected linearly with height, waist measurement, hip measurement, shoulder measurement, and chest measurement. Find the multiple correlation between weights and heights, waist measurements, shoulder measurements, hip measurements, chest measurements, for the data on page 193 if y = weights, x_1 = heights, x_2 = shoulder measurements, x_3 = chest measurements, x_4 = waist measurements, x_5 = hip measurements. Let different members of the class do certain parts of this problem.

4. Assume that y is connected linearly with the 3 variables x_1, x_2, and x_3, by the relation $y = c_0 + c_1x_1 + c_2x_2 + c_3x_3$, and that the following information is given concerning the various means, standard deviations, and correlation coefficients:

$$\begin{aligned}
M_y &= 140.2, & \sigma_y &= 18.1, & r_{yx_1} &= 0.71, & r_{x_1x_3} &= 0.47,\\
M_{x_1} &= 67.6, & \sigma_{x_1} &= 2.5, & r_{yx_2} &= 0.80, & r_{x_2x_3} &= 0.76,\\
M_{x_2} &= 36.8, & \sigma_{x_2} &= 1.9, & r_{yx_3} &= 0.92, & &\\
M_{x_3} &= 20.1, & \sigma_{x_3} &= 1.3, & r_{x_1x_2} &= 0.54. & &
\end{aligned}$$

Find the D's, the predicting equation, and the standard error of prediction. Find the multiple correlation between y and the other variables.

DATA* FOR PROBLEM 3, PAGE 192

Weight y	Height x	Shoulder z	Chest w	Waist u	Hip s	Weight y	Height x	Shoulder z	Chest w	Waist u	Hip s
142	67 5	16 9	35 3	30 2	36 5	144	68 5	17 5	38 0	28 5	36 5
151	73 0	16 5	36 2	29 0	37 3	124	68 2	16 2	33 3	25 8	35.0
130	67 5	17 2	35 0	25 5	35 0	143	69 7	16 4	33 7	28 2	35 3
127	67 0	16 5	36 5	28 1	34 3	122	64 2	16 2	34 7	26 4	33 5
143	65 9	17 2	37 5	29 5	36 0	140	67 5	16 4	36 6	28 5	34 8
132	66 8	16 0	34 2	28 5	35 0	150	66 0	16 0	36 2	30 1	37 2
155	69 5	16 9	35 0	28 5	37 3	141	67 0	16 8	36 0	28 0	36 0
113	66 1	15 5	31 8	25 0	33 0	102	67 6	15 4	32 7	26 2	32 2
161	67 7	17 0	37 6	30 6	38 6	134	67 1	16 4	34 5	27 7	34 2
125	68 5	16 1	32.3	27 2	34 0	136	66 7	15 9	34 2	28 2	35 8
136	68 0	16 0	34 3	27 7	34 2	143	67 6	17 8	36 3	27 2	36 2
136	68 0	17 0	34 2	28 7	35 0	148	68 0	16 0	35 0	28 2	36 2
130	64 2	17 5	34 5	26 5	35 5	131	69 7	15 9	32 7	26 2	34 1
132	63 3	16 1	35 7	28 0	34 2	127	67 5	14 7	32 7	25 2	35 3
122	65 0	17 3	34 0	27 6	34 5	138	66 7	16 4	35 9	28 2	35 5
142	70 5	16 5	34 3	28 5	36 5	140	69 4	15 4	33 6	25 7	35 2
138	70 0	17 0	38 0	27 0	34 0	117	63 5	15 4	34 8	27 1	34 5
129	67 0	17 0	36 5	27 1	34 0	114	67 9	14 5	31 8	24 2	32 2
132	67 4	16 5	35 8	28 6	34 7	142	67 6	16 2	35 4	27 2	36 2
148	72 7	16 6	34 0	27 2	37 3	128	69 7	15 7	34 4	27 6	33 3
139	70 2	16 4	35 2	27 8	35 0	150	67 6	16 8	36 0	29 2	36 2
108	62 4	16 5	30 5	26 0	32 5	132	64 2	15 0	34 7	29 0	34 9
142	67 0	16 7	37 6	29 3	35 7	150	70 8	17 2	35 8	27 3	36 0
135	69 3	17 0	34 0	27 8	35 1	117	68 8	15 7	32 8	25 4	33 2
145	67 3	16 7	35 1	30 0	35 3	138	67 2	17 4	36 5	29 9	35 9
152	71 0	17 3	37 3	29 5	36 6	138	64 4	17 4	35 3	29 5	35 3
132	65 9	16 5	36 7	28 0	34 6	139	67 8	16 4	35 8	28 2	34 2
172	65 1	16 4	38 3	35 0	39 2	122	65 7	15 0	34 7	28 2	33 0
133	67 5	15 5	36 0	28 5	35 5	144	66 4	17 2	36 2	28 4	35 7
144	70 3	16 7	34 5	26 5	36 0	155	71 4	17 2	36 8	27 7	37 3
155	68 3	16 9	36 0	27 7	36 4	125	61 2	14 7	32 0	30 0	34 3
124	69 6	15 3	33 5	25 2	34 7	138	70 0	16 0	34 8	28 0	34 9
141	69 2	15 7	33 0	25 8	33 2	113	66 5	15 4	33 8	24 0	33 3
129	63 0	16 5	35 2	28 3	34 5	144	67 1	16 0	35 8	28 8	36 3
124	64 1	17 2	36 2	28 0	36 8	137	68 2	15 8	34 8	28 6	34.0
115	66 3	15 3	32 1	25 2	32 3	140	67 8	16 2	34 5	27 2	36 2
142	68 6	16 2	35 0	27 6	36 0	143	67 8	16 4	34 8	28 6	35 2
137	64 4	16 5	35 3	26 5	32 0	128	68 3	16 4	34 3	27 3	33 7
144	71 0	16 4	34 0	29 3	36 4	136	68 1	16 2	34 6	26 2	35 0
146	67 0	17 6	34 6	28 0	37 0	152	67 4	16 6	37 0	28 0	37 0
167	67 7	17 2	39 3	30 5	39 5	122	66 6	15 3	32 7	25 1	33 2
140	66 5	17 4	38 9	28 3	35 7	130	65 6	14 9	33 9	28 3	36 1
146	69 5	17 3	35 5	29 1	36 5	109	62 5	15 2	31 4	23 3	33 6
125	69 6	16 0	32 8	27 0	33 7	147	68 6	16 6	37 0	30 4	35 3
129	68 5	16 2	32 7	27 3	35 8	150	71 7	15 9	34 0	28 0	36 3
157	71 0	16 3	35 8	30 2	38 0	159	69 3	16 4	37 0	28 8	36 8
116	63 0	16 3	32 1	26 5	32 5	136	68 7	17 0	36 5	27 8	36 2
152	70 2	16 8	37 2	29 9	35 7	136	68 3	15 2	34 2	28 0	35 0
146	68 2	15 8	36 2	30 0	36 8	172	69 1	17.2	36 8	28 1	36 4
132	67 0	16 1	36 1	29 2	35 0	143	70 7	16 9	35 8	28.3	36.3
148	62 4	16 0	37 2	31 8	37 5	121	60 2	15 9	33 9	28 2	34 2
130	65 3	15 6	34 2	28 2	34 2	159	71 1	16 3	36 8	28 8	35 8
159	71 8	16 8	37 8	29 8	37 8	170	72.0	17.5	39 0	30 3	37.2
145	67 0	15 4	34 6	28 2	35 2	133	63 1	16 2	36 1	28 2	34 2
131	69 2	15 7	34 8	26 5	33 8	144	70 1	17 7	36 0	27 0	35 0
140	66 6	16 2	35 3	28 1	36 1	138	69 4	17 1	35 6	26 2	34 7
153	70 1	16 0	35 7	29 7	35 7	163	70 7	18 4	38 6	30 0	37 7
119	61 9	15 9	34 8	29 2	33 8	169	72 2	17 7	38 2	30 0	38 3
155	68 0	16 3	37 0	29 2	38 2	121	69 0	15 7	34 5	27 5	33 7
150	70 0	16 5	35 3	29.2	36 8	140	66 4	16 2	35 8	35 3	20.2

* These data were taken from "Synopsis of Elementary Mathematical Statistics," by H. C. Carver.

PARTIAL CORRELATION COEFFICIENT

When x and y are connected linearly by the relation

$$y = a + bx,$$

values of y are predicted from values of x. By making the sum of the squares of the residual errors a minimum, the value of b is found to be $b = \Sigma\bar{x}\bar{y}/\Sigma\bar{x}^2$. When x and y are connected linearly the x values are predicted from values of y by the relation

$$x = c + dy.$$

By making the sum of the squares of the residual errors a minimum, the value of d is found to be $d = \Sigma\bar{x}\bar{y}/\Sigma\bar{y}^2$.
The correlation coefficient between x and y is

$$r_{yx} = \frac{\Sigma\bar{x}\bar{y}}{\sqrt{\Sigma\bar{x}^2\Sigma\bar{y}^2}}.$$

This same quantity can be found by taking the square root of the product of b and d, for

$$r_{yx} = \sqrt{bd} = \sqrt{\frac{(\Sigma\bar{x}\bar{y})^2}{\Sigma\bar{x}^2\Sigma\bar{y}^2}} = \frac{\Sigma\bar{x}\bar{y}}{\sqrt{\Sigma\bar{x}^2\Sigma\bar{y}^2}}.^*$$

This shows that the correlation coefficient between x and y is the square root of the product of the coefficient of x in the predicting equation for y and the coefficient of y in the predicting equation for x, or $r_{yx} = \sqrt{db}$.

Consider the three variables x, y, and z connected by the linear relation

$$y = a + bx + cz, \quad \text{or} \quad \bar{y} = b\bar{x} + c\bar{z}.$$

It is also assumed that any 2 of these 3 variables are connected linearly.
It is assumed that there are n "corresponding" values of x, y, and z given from observations. From the observational equations the following normal equations are obtained:

$$b\Sigma\bar{x}^2 + c\Sigma\bar{z}\bar{x} = \Sigma\bar{x}\bar{y}$$

$$b\Sigma\bar{x}\bar{z} + c\Sigma\bar{z}^2 = \Sigma\bar{z}\bar{y},$$

* r_{yx} is the geometric mean of b and d.

from which

$$b = \frac{\Sigma\bar{z}^2\Sigma\bar{x}\bar{y} - \Sigma\bar{x}\bar{z}\Sigma\bar{z}\bar{y}}{\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{x}\bar{z})^2}.$$

Now consider that x is predicted from y and z by the equation

$$\bar{x} = d\bar{y} + g\bar{z}.$$

The normal equations are

$$d\Sigma\bar{y}^2 + g\Sigma\bar{z}\bar{y} = \Sigma\bar{x}\bar{y},$$

$$d\Sigma\bar{y}\bar{z} + g\Sigma\bar{z}^2 = \Sigma\bar{x}\bar{z},$$

from which

$$d = \frac{\Sigma\bar{z}^2\Sigma\bar{x}\bar{y} - \Sigma\bar{z}\bar{y}\Sigma\bar{x}\bar{z}}{\Sigma\bar{y}^2\Sigma\bar{z}^2 - (\Sigma\bar{y}\bar{z})^2}.$$

The square root of the product of bd is

$$\sqrt{bd} = \sqrt{\frac{(\Sigma\bar{z}^2)^2(\Sigma\bar{x}\bar{y})^2 - 2\Sigma\bar{z}^2\Sigma\bar{x}\bar{y}\Sigma\bar{z}\bar{y}\Sigma\bar{x}\bar{z} + (\Sigma\bar{x}\bar{z})^2(\Sigma\bar{z}\bar{y})^2}{[\Sigma\bar{x}^2\Sigma\bar{z}^2 - (\Sigma\bar{x}\bar{z})^2][\Sigma\bar{z}^2\Sigma\bar{y}^2 - (\Sigma\bar{y}\bar{z})^2]}}$$

$$= \sqrt{\frac{r_{xy}{}^2 - 2r_{xy}r_{xz}r_{yz} + r_{xz}{}^2r_{zy}{}^2}{(1 - r_{xz}{}^2)(1 - r_{zy}{}^2)}}.$$

The partial correlation coefficient is defined to be

$$r_{yx\cdot z} = \sqrt{bd} = \frac{r_{xy} - r_{xz}r_{zy}}{\sqrt{(1 - r_{xz}{}^2)(1 - r_{zy}{}^2)}}. \tag{10.20}$$

The above correlation coefficient was obtained as in finding r_{yx}, that is, by finding $\sqrt{bd}$. The values of c and g were not used. This amounts to holding z constant. This correlation is called partial because it is the correlation obtained when the third variable is held constant. Suppose height is predicted from weight and chest measurements, then the partial correlation coefficient

$$r_{\text{height weight}\cdot\text{chest measurement}}$$

is the correlation between height and weight when chest measurement is held constant, that is, when people are considered who have the same chest measurement. The Pearson linear correlation coefficient between heights and weights for people with 37-inch chest measurement is what is meant by holding chest measurement constant. The correlation should be about the same if people

were considered with chest 38 inches or 40 inches. Instead of separating all the people in the distribution considered with a certain chest measurement and then finding the ordinary correlation coefficient between heights and weights, it is only necessary to find the partial correlation coefficient, for it does the same thing.

To be able to use the partial correlation coefficient it must be ascertained whether or not the variables are linearly connected in all possible pairs, and in every other way. This should be done, but it is very seldom done. The average person uses the partial correlation coefficient without determining whether or not the variables are connected linearly in all possible pairs.

EXAMPLE. Consider that the four-year average of a university student is a measure of his general ability For a group of students at a certain university the following correlations were obtained, where m = freshman mathematics average, s = science average, and a = four-year average: $r_{ms} = 0.54$, $r_{ma} = 0.80$, $r_{sa} = 0.57$.

The correlation between freshman mathematics and science with general ability eliminated is

$$r_{ms.a} = \frac{r_{ms} - r_{ma}\, r_{as}}{\sqrt{(1 - r_{as}^2)\,(1 - r_{ma}^2)}} = 0.17,$$

which seems to indicate that there is not very much correlation between freshman mathematics average and science average.

PROBLEMS

1. Find the correlation coefficient between freshman mathematics averages and education averages with general ability eliminated, if m = mathematics average, e = education average, a = four-year average, or general ability, and $r_{me} = 0.34$, $r_{ma} = 0.66$, and $r_{ea} = 0.54$. Discuss your result for poor students, for average students, and for the best students.

2. Find the correlation between weights and thigh measurements with heights eliminated for the data on pages 151 and 161.

3. Find the correlation coefficient between freshman mathematics averages and freshman English averages with general ability eliminated for data on page 188.

4. Discuss formula (10.20) when $r^2_{yx \cdot z} = 1$.

5. Discuss this same formula when $r_{xz}^2 = 1$.

6. Discuss this when $r_{xy}^2 = 1$.

7. Find the partial correlation coefficient between weights and waist

measurements, with shoulder measurements held constant, for the data on page 193.

8. Use data on page 193 for finding the correlation between weights and waist measurements for those whose shoulder measurements were between 15.5 inches and 16.0 inches inclusive. How does this compare with the result of problem 7?

9. Use data on page 193 for finding the correlation coefficient between weights and waist measurements for those whose shoulder measurements were between 16.5 inches and 17.0 inches inclusive. Compare this with results of problems 7 and 8.

10. By using formula (10.16) and formula (10.20) prove that

$$(10.21) \qquad 1 - (r_{yx\cdot z}{}^2) = \frac{1 - (r_{y\cdot xz})^2}{1 - r_{yz}{}^2}.$$

11. Use this formula to find the partial correlation coefficient between mathematics grades and English grades with general ability eliminated for data on page 188 and the result of problem 5 on page 188.

12. If $\bar{y} = b\bar{x} + c\bar{z}$, prove that this equation can be written in the form

$$[\sqrt{(1 - r_{xz}{}^2)}]\bar{y} = \frac{\sigma_y}{\sigma_x}(\sqrt{1 - r_{yz}{}^2})r_{yx\cdot z}\cdot\bar{x} + \frac{\sigma_y}{\sigma_z}(\sqrt{1 - r_{xy}{}^2})r_{yz\cdot x}\cdot\bar{z}\cdot$$

13. If $r_{yz} = \pm 0.80$, plot relation (10.21), after reducing to the simplest form. Discuss the results.

14. If the multiple correlation coefficient between y and (x and z) is equal to 1, what is the partial correlation coefficient between y and x with z held constant?

15. If $r_{yx\,z} = +0.80$, plot relation (10.21). Discuss.

16. If $\bar{y} = b\bar{x} + c\bar{z} + d\bar{w}$, find an expression for the partial correlation coefficient between y and x with z and w held constant, or find an expression for $r_{yx\,zw}$ in terms of the fundamental summations and also in terms of r's.

17. Prove that

$$(10.22).\; r_{yx\cdot zw} = \frac{r_{yx\,z} - r_{yw\,z}r_{xw\cdot z}}{\sqrt{(1 - r_{yw.z}{}^2)(1 - r_{xw.z}{}^2)}} = \frac{r_{xy.w} - r_{yz.w}\cdot r_{xz\,w}}{\sqrt{[1 - r_{yz.w}{}^2]\,[1 - r_{xz\,w}{}^2]}}.$$

18. Use formula (10.22) and the data on page 193 to find the partial correlation between weight and hip measurement with height and shoulder measurement held constant.

TETRACHORIC CORRELATION

This section presents a correlation coefficient between variables which cannot be measured or which are not quantitative; for example, blindness, deafness, dispositions, mealiness of apples, desirableness of potatoes, etc. Karl Pearson derived a formula for finding

a measure of the correlation between two qualitative variables.* Consider the following example concerning inheritance of eye-color in man.

Grandmothers

	Tint	Gray or Lighter	Dark Gray or Darker	Totals
Granddaughters	Gray or Lighter	a 254	b 136	$a+b$ 390
	Dark Gray or Darker	c 156	d 193	$c+d$ 349
	Totals	$a+c$ 410	$b+d$ 329	N 739

The above table gives the number of granddaughters with gray or lighter eyes whose grandmothers had eyes which were gray or lighter and those which had eyes that were dark gray or darker, and also the number of granddaughters with dark gray or darker eyes whose grandmothers had eyes which were gray or lighter and those with eyes which were dark gray or darker. Pearson's formulas enable one to find the correlation coefficient between eye-color of granddaughter and eye-color of grandmother. These formulas are as follows:

(10.23) $\dfrac{c+d}{N}$ = the area under the normal curve between k and ∞,

(10.24) $\dfrac{b+d}{N}$ = the area under the normal curve between h and ∞,

(10.25) $H = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{h^2}{2}}$, or the height of the normal curve at h,

(10.26) $K = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{k^2}{2}}$, or the height of the normal curve at k,

(10.27) $hkr^2 + 2r = \dfrac{2(ad - bc)}{H \cdot K \cdot N^2}$,

* Karl Pearson, "Mathematical Contributions to the Theory of Evolution—VII. On the Correlation of Characters Not Quantitatively Measurable," *Phil. Trans.*, 195–A, pages 1–47.

where h and k can be found from the area table for the normal curve, and H and K can be found from the ordinate table of the normal curve The quantity r is found by solving the quadratic equation (10 27), and is the coefficient of correlation sought. For the above table pertaining to eye-color we have

$$\frac{c+d}{N} = \frac{349}{739} = 0.4723 = \text{the area under the normal curve between } 0.0696 \text{ and } \infty\text{; from Table I } (k = 0.06961)$$

$$\frac{b+d}{N} = \frac{329}{739} = 0.4452 = \text{the area under the normal curve between } 0.1381 \text{ and } \infty\text{; from Table I } (h = 0.1381)$$

$$H = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(0.1381)^2} = 0.3952, \text{ from Table II;}$$

$$K = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(0.0696)^2} = 0.398, \text{ from Table II.}$$

The quadratic equation (10.27) becomes

$$0.009612r^2 + 2r = \frac{55{,}612}{85{,}899{,}154} = 0.6475,$$

from which

$$r = 0.3225^+,$$

which is the correlation coefficient between eye-color of grandmother and granddaughter.

PROBLEMS

1. Find the tetrachoric correlation coefficient between blindness and deafness for the following data taken from the 1930 census of the United States.

Deafness \ Blindness	Present	Absent	Totals
Present	1,942	55,142	57,084
Not Present	61,747	122,656,215	122,717,962
Totals	63,689	122,711,357	122,775,046

2. Find the tetrachoric correlation coefficient between vaccination and recoveries from the following data.

	Deaths	Recoveries
Vaccinated	546	2,602
Unvaccinated	130	100

3. Find the tetrachoric correlation coefficient between toughness of skin and mealiness of potatoes of samples taken from stores at random.

		Skin	
		Tough	Not tough
Mealiness	Mealy	12	93
	Not Mealy	30	14

CHAPTER 11

SAMPLING

AVERAGE OF SAMPLE AVERAGES

The most important part of statistics is sampling from an unknown parent population. Often all the information concerning the parent population is obtained from a rather small sample or from a small number of samples. In physics, many times, 10 or 12 experiments are made to discover the nature or law of a certain phenomenon. The botanist picks at random a sample of a certain plant, investigates the characteristics of this sample, and assigns to this species the information obtained from the sample. Business forecasting depends on information obtained from samples; in fact, almost every empirical formula is obtained from sampling data.

Since sampling is so important, a few of the basic ideas concerning sampling will be presented in this chapter. These ideas will be presented first from the study of sampling from a known parent population and later from an unknown parent One of the first things to investigate is the nature of the distribution made up of averages of the samples drawn from a parent.

An example will be given to introduce this phase of sampling. Consider the following parent distribution:

v	$f(v)$
0	1
1	2
2	9
3	28
4	66
5	121
6	175
7	197
8	175
9	121
10	66
11	28
12	9
13	2
14	1
	1,001

The mean, standard deviation and skewness are as follows:

$$M_v = 7.0000,$$
$$\sigma_v = 2.002,$$
$$\alpha_{3 \cdot v} = 0.0000.$$

From this parent distribution 400 samples, each of 200 variates, were drawn at random. Cards with the variates written thereon were put in a bag, from which drawings were made. The average of each sample of 200 was obtained, giving 400 values. These 400 values formed the frequency distribution given below, which is the distribution of the means of these 400 samples.

Classes z	$f(z)$
6.625–6.674	0
6.675–6.724	2
6.725–6.774	12
6.775–6.824	20
6.825–6.874	22
6.875–6.924	46
6.925–6.974	57
6.975–7.024	66
7.025–7.074	55
7.075–7.124	50
7.125–7.174	40
7.175–7.224	15
7.225–7.274	11
7.275–7.324	3
7.325–7.374	1
	400

The mean, standard deviation, and skewness of the distribution of the means are as follows:

$$M_z = 7.0055,$$
$$\sigma_z = 0.121,$$
$$\alpha_{3:z} = 0.0001.$$

The mean of the parent is 7.0000, and the mean of the distribution of sample means is 7.0055. These values are nearly the same. The standard deviation of the parent distribution is 2.002; the standard deviation of the distribution of sample averages is 0.121. (These distributions were taken from "An Application of Thiele's Semi-Invariants to the Sampling Problem," by C. C. Craig, *Metron*, Vol. 7, No. 4, 1928, pages 3–75.) The smallness of the standard deviation of the distribution of the sample averages shows that all the sample averages are very near their mean and the mean of the parent.

The range of the variates in the parent is 15, while the range of the distribution of sample means is only 0.75, which again shows how near the variates in the distribution of sample averages lie with reference to the mean of the sample averages and to the mean of the parent. Compare the σ's of the two distributions. In this example all the sample averages were within a distance of 0.375 from the mean of the parent population. From the theory

of probability it is almost impossible to get a sample average, where the sample consists of 200 variates, to differ from the mean of the parent by as much as 0.5, hence this example shows that as far as the mean is concerned a mean of a good size sample is about as good as the mean of the parent. When the parent is unknown, the mean of a sample, or of a few samples, is about as good as we can get without going to a great deal of trouble to secure other samples or a larger sample. The two distributions on pages 201 and 202 are plotted in Fig. 11.1 in percentages. Notice the ranges of the distributions and the locations of the larger frequencies.

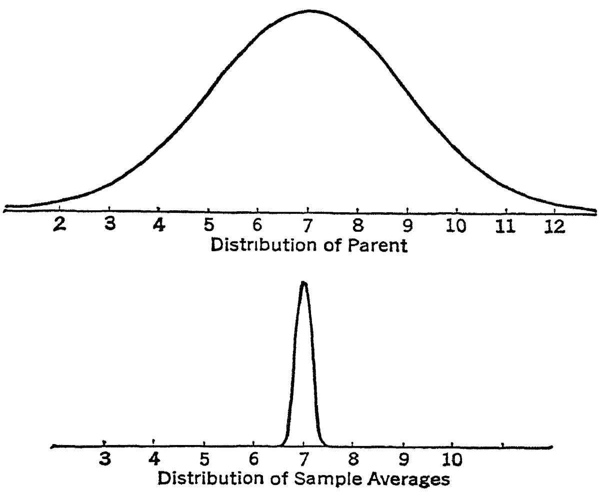


FIG. 11.1

Let us now consider the general case. Out of a parent population consisting of s variates, $v_1, v_2, v_3, \ldots, v_s$, are taken at random samples of r variates. Each sample average is obtained, and then the arithmetic mean of all possible sample averages is found. The parent population may consist of the heights of soldiers in one particular camp, or the lung capacities of students in a certain college, or the numbers of bushels of wheat grown per year in a certain state for the last 50 years, or the numbers of petals on a certain kind of flower, etc. Let the sample averages be represented by z, where z_1 is the first sample average, z_2 the second, etc. Let the mean, M_v, the standard deviation, σ_v, and the skewness, $\alpha_{3:v}$, be known of the parent. The sample averages are:

$$
\begin{aligned}
z_1 &= \frac{v_1 + v_2 + \ldots v_r}{r}, \\
z_2 &= \frac{v_1 + v_2 + \ldots + v_{r-1} + v_{r+1}}{r}, \\
\ldots &= \ldots\ldots\ldots \ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
z_{{}_sC_r} &= \frac{v_{s-r+1} + v_{x-r+2} + \ldots + v_{s-1} + v_s}{r}
\end{aligned}
$$

Adding

$$M_z = \frac{\Sigma z_i}{{}_sC_r} = \frac{r \cdot {}_sC_r \sum_1^s v_i}{r \cdot {}_sC_r \cdot s} = \frac{\sum_1^s v_i}{s} = M_v,$$

where M_z is the mean of all possible sample means; no sample is taken twice.

The above summation was obtained by adding all the variates in every row. There are ${}_sC_r$ rows or possible samples of r variates which can be formed from the s variates in the parent. Hence the sum of all the variates in the ${}_sC_r$ samples is $r \cdot {}_sC_r$ as there are r variates in each of the ${}_sC_r$ rows. Any variate, say v_3, appears in this sum as often as any other variate. Since there are s variates in the parent and since each variate appears as often as any other, we divide the above sum by s and let the summation run from 1 to s. Dividing by s gives the number of times each variate appears in the sum. After reducing the above average of the z's we find that the mean of all possible sample averages is equal to the mean of the parent population, or that

$$M_z = M_v. \tag{11.1}$$

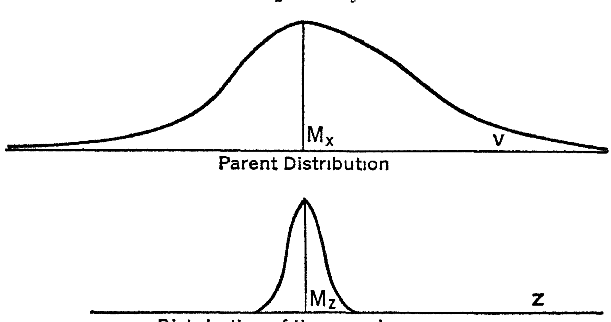


FIG. 11.2

Let the frequency curves in Fig. (11.2) represent the parent distribution and the distribution made up of all possible sample averages.

The lower distribution represents frequencies of the sample averages which are plotted on a z axis about the mean of the z's which is equal to the mean of the parent. The range of the distribution made up of all possible sample averages is much less than the range of the parent.

PROBLEMS

1. Let the parent population consist of the variates $v_1, v_2, \ldots, v_{10}$. Write all possible sample averages of 3 in the sample where no sample is repeated. Find the average of these sample averages. Does this check with (11.1)?

2. How many possible samples of 6 variates to the sample can be taken from a parent of 10 variates, where repetitions of samples are not allowed? With 4 variates in the samples? Five in the sample?

3. If there are 1,000 students whose average weight is 140 pounds and if 500 students are taken at random from this 1,000, what would you expect the average weight of these 500 to be? What would you expect if 100 were drawn at random? Would the standard deviation of the distribution of the sample means help you to answer this question? Explain.

4. The following data give the right thigh measurements in inches of

20	22	20	17	17	19	20	21	21	22
19	20	21	18	20	21	19	22	19	20
23	20	20	17	22	20	19	19	23	18
20	24	23	18	21	21	19	19	26	20
20	21	21	20	20	19	22	19	18	20
22	21	20	19	21	19	20	22	22	19
18	20	21	17	23	20	21	20	25	24
21	19	18	18	17	20	19	24	21	18
18	18	19	19	20	19	23	19	23	18
19	20	19	21	21	20	19	19	20	20
21	22	21	19	19	18	19	23	24	23
20	19	21	20	18	17	22	18	20	20
17	19	20	21	20	21	23	18	20	19
23	22	24	22	23	18	20	19	23	20
20	20	24	19	20	19	20	22	21	21
19	19	19	20	23	18	21	19	20	21
23	22	19	19	20	18	23	21	26	19
22	19	23	20	19	18	21	19	21	19
18	26	20	18	23	21	22	18	20	21
19	21	21	22	20	21	21	20	19	20

men of 19 years of age. Have each member of the class pick 50 measurements and find the average of this sample of 50 measurements. Examine these sample averages with reference to the average of the parent, which is $M_v = 20.28$. Find the average of all sample averages obtained by members of the class, and compare this with the mean of the parent. Find some scheme where each member will secure a sample which is different from the others. For example, let one member take every other one until 50 are obtained, another take every fifth one until 50 are obtained, etc.

STANDARD DEVIATION OF THE DISTRIBUTION OF SAMPLE AVERAGES

As before, let $\bar{v}_i = v_i - M_v$, which is the deviation of variate v_i from the mean of the v's or the mean of the parent. Let the deviation of the first sample average from the mean of all sample averages be

$$(11.2) \quad \bar{z}_1 = z_1 - M_z = z_1 - M_v = \frac{v_1 + v_2 + \ldots + v_r}{r} - M_v$$

$$= \frac{(v_1 - M_v) + (v_2 - M_v) + \ldots + (v_r - M_v)}{r}$$

$$= \frac{\bar{v}_1 + \bar{v}_2 + \ldots + \bar{v}_r}{r},$$

which expresses the deviation of the first sample average from the mean of all possible sample averages in terms of the deviations of the variates in the parent from the mean of the parent. The standard deviation of the distribution of all possible sample averages is the square root of the average of the $\bar{z}^2$'s. These deviations squared are:

$$\bar{z}_1^2 = \left[\frac{\bar{v}_1 + \bar{v}_2 + \ldots + \bar{v}_r}{r}\right]^2 = \frac{\Sigma\bar{v}_i^2 + 2\Sigma\bar{v}_i\bar{v}_j}{r^2}, \text{ where } i \neq j.$$

$$\bar{z}_2^2 = \left[\frac{\bar{v}_1 + \bar{v}_2 + \ldots + \bar{v}_{r-1} + \bar{v}_{r+1}}{r}\right]$$

$$= \frac{\Sigma\bar{v}_i^2 + 2\Sigma\bar{v}_i\bar{v}_j}{r^2}, \; i \neq j, i, j \neq r$$

$$\ldots = \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots,$$

$$\bar{z}^2_{sC_r} = \left[\frac{\bar{v}_{s-r+1} + \ldots + \bar{v}_{s-1} + \bar{v}_s}{r}\right]^2 = \frac{\Sigma\bar{v}_i^2 + 2\Sigma\bar{v}_i\bar{v}_j}{r^2}; i, j > s - r$$

$$\mu_{2\ z} = \frac{1}{r^2} \cdot \frac{r_sC_r\overset{s}{\Sigma}\bar{v}_i^2}{s \cdot {}_sC_r} + \frac{2_sC_r \cdot {}_rC_2\overset{s}{\Sigma}\bar{v}_i\bar{v}_j}{{}_sC_r \cdot {}_sC_2} = \frac{1}{r}\left[\mu_{2\ v} + \frac{2(r-1)}{s(s-1)}\sum_1^s \bar{v}_i\bar{v}_j\right],$$

where $i \neq j$.

The sum of the square terms was found by adding the squares of the variates in each row. Since there are r squares in each row and ${}_sC_r$ rows, there are $r \cdot {}_sC_r$ variates squared in all the rows.

The average of these square terms is

$$r \cdot {}_sC_r \sum_{i=1}^{s} \bar{v}_i^2 / s \cdot {}_sC_r,$$

where the summation runs from 1 to s. Since each variate squared appears as often as any other and since there are s variates, then putting an s in the denominator gives the number of times each variate squared appears.

The sum of the product terms is found in a similar manner. There are ${}_rC_2$ product terms in each row and ${}_sC_r$ rows, hence there are ${}_rC_2 \cdot {}_sC_r$ product terms in all rows. Each product term like v_3v_7 appears as often as any other product term. As there are ${}_sC_2$ possible product terms the average of the product terms is equal to

$$\frac{2 \cdot {}_rC_2 \cdot {}_sC_r\overset{s}{\Sigma}\bar{v}_i\bar{v}_j}{r^2 \cdot {}_sC_r \cdot {}_sC_2}, i \neq j$$

Dividing by ${}_sC_2$ gives the number of times each product term is repeated.

The summation, $\overset{s}{\Sigma}\bar{v}_i\bar{v}_j$, $i \neq j$; can be expressed in terms of $\Sigma\bar{v}_i^2$. The sum of the deviations of the variates from the mean of the variates is zero as was proven in an earlier chapter, hence

$$0 = (\bar{v}_1 + \bar{v}_2 + \ldots + \bar{v}_s)^2 = \Sigma\bar{v}_i^2 + 2\Sigma\bar{v}_i\bar{v}_j,\ i \neq j,$$

or

$$(11.3) \qquad 2\Sigma\bar{v}_i\bar{v}_j = -\ \Sigma\bar{v}^2.$$

Substituting this in the expression for $\mu_{2\,z}$, gives

$$\mu_{2:z} = \frac{1}{r}\left[\mu_{2\,v} - \frac{r-1}{s-1}\mu_{2.v}\right].$$

Therefore

$$\sigma_z = \sqrt{\frac{s-r}{r(s-1)}} \cdot \sigma_v. \tag{11.4}$$

This formula shows that the standard deviation of all possible sample averages is equal to $\sqrt{(s-r)/r(s-1)}$ times the standard deviation of the parent. This gives the standard deviation of the distribution of sample averages in terms of s, the number of variates in the parent; r, the number of variates in each sample; and σ_v, the standard deviation of the parent. When the mean and standard deviation of the parent are known, the mean and standard deviation of the distribution of sample averages can be obtained.

PROBLEMS

1. Given a parent population of 5 variates, v_1, v_2, v_3, v_4, v_5. Set up all possible sample averages of 3 variates, find the $\bar{z}$'s and the standard deviation of all possible sample averages, where repetitions are not allowed.

2. Given a parent population consisting of the heights of 1,000 men, with mean 67.6 inches, and standard deviation equal to 2.5 inches. Find the standard deviation of all possible sample averages if each sample contains 200 men. If each sample contains 500 men. If each sample contains 800 men.

3. Plot $\sigma_z = \sqrt{(s-r)/r(s-1)} \cdot \sigma_v$, where $s = 1{,}000$, $\sigma_v = 2.5$. Discuss the plot. How large must r be so that σ_v is less than one-half of σ_v? Less than one-tenth as large?

4. If all the sample averages are within 3.2 standard deviations of the mean of all possible sample averages, what are the ranges for the sample averages in the parts of problem 2?

SKEWNESS OF THE DISTRIBUTION OF THE MEANS

It can be shown that the skewness of the distribution made up of all possible sample means is

$$\alpha_{3:z} = \frac{s-2r}{s-2}\sqrt{\frac{s-1}{r(s-r)}}\,\alpha_{3:v}. \tag{11.5}$$

If s approaches infinity, formulas (11.4) and (11.5) become

$$\sigma_z = \sigma_x/\sqrt{r}; \tag{11.6}$$

$$\alpha_{3:z} = \alpha_{3:x}/\sqrt{r}, \tag{11.7}$$

as can be seen by dividing both numerator and denominator in the formulas for a finite parent by s and then allowing s to become infinitely large.

Formula (11.7) indicates that when r is large the skewness of the distribution of the means is nearly equal to zero; this means that the distribution of the sample means is approximately symmetrical for large values of r.

Example. There are 4,000 students on a college campus. The average weight of these students is 145 pounds; the standard deviation is 20 pounds; the skewness of the distribution made up of the weights of the students is 0.3. If a sample of 1,000 students is taken at random from the student body, what is the probability that the average weight shall be greater than 148 pounds?

$$Z_1 = \frac{x_1 + x_2 + \ldots + x_{1,000}}{1,000}.$$

$M_z = M_x = 145$ pounds, which is the mean of all possible averages of 1,000 weights. The standard deviation and skewness are:

$$\sigma_z = 20\sqrt{\frac{s - r}{r(s - 1)}} = 20\sqrt{\frac{3,000}{1,000(3,999)}} = 0.55 \text{ lb.}$$

$$\alpha_{3:z} = 0.006; \text{ call it zero.}$$

$$t = \frac{148 - 145}{0.55} = 5.5 \text{ standard units,}$$

which means that 148 pounds, the average of 1 sample taken at random, deviates from the average of all possible sample averages by 5.5 standard units. On examining normal probability tables it is seen that the probability that a sample average as large as or larger than 148 is very small indeed. Very few tables record any area beyond 5.5 standard units. The difference between 145 and 148 is only 3 pounds, yet it is almost impossible to take at random a sample of 1,000 students from the original 4,000 and secure an average which differs by as little as 3 pounds from the mean of the entire group. The work so far shows that the mean of a representative sample differs by a very small amount from the mean of the parent, when the sample is large. The mean is usually written as $M_z \pm \sigma_{Mz}$.

PROBLEMS

1. The mean, standard deviation, and skewness of a parent population with 1,000 variates are respectively 36.8 inches, 1.2 inches, and 0.6. Find the mean, standard deviation, and skewness of the distribution made up of all possible sample means of 400 to the sample. When the sample contains 500 variates. When the sample contains 800 variates.

2. Consider a parent with an infinite number of variates, where $M_x = 40.3$ gallons, $\sigma_x = 1.7$ gallons, and $\alpha_{3 \cdot x} = 0.4$. Find the mean, standard deviation, and skewness of the distribution made up of all possible sample averages, when the samples contain 1,000 variates. When the samples contain 100 variates, 10,000 variates.

3. There are 1,200 first-year men on the campus. The following is known concerning the distribution of their heights: $M_v = 68.1$ inches, $\sigma_v = 2.3$ inches, $\alpha_{3 \cdot v} = 0.2$ Two hundred first-year students are picked at random; find the probability that the average of these heights will be: (*a*) greater than 68.7 inches; (*b*) less than 67.8 inches; (*c*) greater than 67.6 inches and less than 68.7 inches if measurements are made to the nearest 0.1 inch.

4. Out of 400 head of cattle 200 are picked. It is found that the average number of pounds of milk for the first group is 39, for the second group is 36; also that the standard deviations are respectively 4.5 and 5.4 pounds. Is the difference due to random sampling? (*Hint.* Find the mean and standard deviation of the pounds of milk from the 400 cows.)

5. Solve problem 4 when the first group contains 100 cows and the second contains 300.

6. If the standard deviation of all possible samples of 1,000 from a parent population consisting of s items is 0.55 and the standard deviation of the parent is 19.8, find s the number of items in the parent.

7. Find the standard deviation of all sample averages obtained by the class when solving problem 4 on page 205. Compare this value with the standard deviation of the parent and with the value obtained from formula (11.4) (although this formula is for all possible samples). Compare this value with the standard deviation of the parent divided by $\sqrt{n}$, where n is the number in the sample.

8. Find the skewness of the distribution made up of the sample averages which the class obtained when solving problem 4 on page 205. Compare this with that obtained from formula (11.5).

9. Given a parent population of the variates v_1, v_2, v_3, and v_4. Set up all possible sample averages of 2 variates, and find the skewness of the averages of all possible samples.

10. How many possible samples of 5 variates can be drawn from a parent consisting of 100 variates? How many when there are 25 variates in the sample?

11. A parent contains 10,001 variates, with mean equal to 69.7 feet and standard deviation equal to 7.4 feet. How large must the sample be so that the standard deviation of the mean of the sample averages will be less than 2 feet? Less than 1 foot? Less than $\frac{1}{2}$ foot?

12. A parent population contains an infinite number of variates with mean equal to 1,927 gallons and standard deviation equal to 160 gallons. How large will a sample have to be so that the standard deviation of the distribution made up of all possible sample means will be less than 80 gallons? Less than 40 gallons? Less than 20 gallons? Less than 10 gallons? Less than 5 gallons? Less than 1 gallon?

13. If the skewness of the parent population in problem 12 is 1.1, how large will a sample have to be so that the skewness of the distribution of all possible sample averages will be less than 0 1?

14. What is the mean of all possible sample averages of 10,000 variates taken at random from the parent in problem 12? Of 100,000 variates? Of 10 variates?

15. A parent population contains an infinite number of variates with standard deviation equal to 25 cubic feet. Plot and discuss the curve

$$\sigma_z = \frac{\sigma_v}{\sqrt{r}} = \frac{25}{\sqrt{r}}.$$

16. Assume that the skewness of the parent in problem 15 is 0.8. Plot and discuss the curve

$$\alpha_{3\ z} = \frac{\alpha_{3\ v}}{\sqrt{r}} = \frac{0\ 8}{\sqrt{r}}.$$

SAMPLING FROM AN UNKNOWN PARENT

It was shown, when the parent population was known, that the average of all possible sample averages was the same as the average of the parent. It was also pointed out that, when the sample was large, its mean did not differ very much from the mean of the parent. In sampling, one never takes all possible samples, but one or two good-sized samples, using the average of these to represent the mean of the parent. This average will not differ very much from the mean of the parent if the sample is fairly large.

When the parent is unknown the average of a large sample is also considered to be near the mean of the parent, since the range for the distribution of all possible sample averages is very small. Hence the mean of a good-sized sample or the average of the means of several good-sized samples will give a good approximation of

the mean of the parent. This is usually accepted as a good estimate of the mean of the parent.

It is now necessary to find the standard deviation of all possible sample means, or the standard deviation of the distribution made up of all possible sample averages when the parent is unknown. This standard deviation when found will indicate how accurate the mean we have obtained from one or several samples is. When the parent was known the standard deviation of the sample averages was found on page 208. There the standard deviation of the parent was known and was used to find the standard deviation of the distribution of the means of the samples. When the parent is unknown its standard deviation is unknown, hence this formula cannot be used. It is now necessary to find an estimate for this standard deviation.

The standard deviation of the sample means shall be considered to be the square root of the average of all possible second moments of the samples about their respective means. Consider for the present that the parent of s variates is known and that all possible samples can be obtained. Take samples of r variates from the parent and find the second moment about the mean of this sample. Take another sample and find the second moment about the mean of this sample. Do this for all possible samples. Find the average of all these second moments. The second moment about the mean of the first sample is

$$\mu_{2:x_1} = \frac{\overset{r,1}{\Sigma} v_i^2}{r} - \left(\frac{\Sigma v_i}{r}\right)^2.$$

The following gives the second moments about the means of the respective samples, and the average of all of them

$$\mu_{2:x_1} = \Sigma(v_i - M_{x_1})^2/r = \frac{\Sigma v_i^2}{r} - \left(\frac{\Sigma v_i}{r}\right)^2$$

$$= \frac{1}{r^2}\left[(r-1)\sum^{r,1} v_i^2 - 2\sum^{r,1} v_i v_j\right]^*, \; i \neq j$$

$$\mu_{2:x_2} = \qquad = \frac{1}{r^2}\left[(r-1)\sum^{r,2} v_i^2 - 2\sum^{r,2} v_i v_j\right],$$

. . . .

$$\mu_{2:z_{sC_r}} = \qquad = \frac{1}{r^2}\left[(r-1)\sum^{r,sC_r} v_i^2 - 2\sum^{r,sC_r} v_i v_j\right]$$

$$\frac{\sum_{i=1}^{sCr}\mu_{2:z_i}}{{}_sC_r} = \frac{1}{r^2}\left[\frac{(r-1)r({}_sC_r)\sum_{i=1}^{s} v_i^2}{s({}_sC_r)} - \frac{2({}_rC_2)({}_sC_r)\sum_{i,j}^{s} v_i v_j}{({}_sC_2)({}_sC_r)}\right], i \neq j$$

$$= \frac{r-1}{r}\left[\mu'_{2:v} - \frac{1}{s(s-1)}(s^2\mu'^2_{1:v} - s\mu'_{2:v})\right]; (\Sigma v)^2 = s^2 M_v^2$$

$$= \frac{r-1}{r}\left[\frac{s(\mu'_{2\cdot v} - \mu'^2_{1:v})}{s-1}\right] = \frac{s(r-1)}{r(s-1)}\mu_{2:v}.$$

Thus the average of all possible second moments about the means of the respective sample means is

$$M\mu_{2:z} = \frac{s(r-1)}{r(s-1)}\mu_{2\cdot v}, \tag{11.8}$$

which is not the same as the second moment about the mean of the parent. The quantity $M\mu_{2\cdot z}$ is the average of all second moments of the samples about the respective means of the samples.

Take at random a sample of r variates from a parent population and find the second moment of this sample about its mean; this may be a value which is not far from $M\mu_{2\cdot z}$, and again it may not be. Take several samples, and find the average of their second moments about their respective means. This average will be a better estimate of $M\mu_{2:z}$ than the second moment about the mean of just one sample. The quantity $M\mu_{2\ z}$ will be considered to be the best value for the second moment about the mean of any particular sample or the average of the second moments about the means of several samples. Since $M\mu_{2:z}$ is never known, an estimate of it will be considered to be the second moment about the mean of a good-sized sample. We shall call the second moment about the mean of a sample the square root of $M\mu_{2:z}$, or

$$\sigma_{\text{sample}} = \sqrt{M\mu_{2:z}} = \sigma_{\text{parent}}\sqrt{\frac{s(r-1)}{r(s-1)}}.$$

As a rule, the parent population is not known, and hence σ_{parent} and s are not known. All the information that is known is that

* The r, 1 above Σ means the *first* sample with r variates in the sample; r, 2, the second, etc.

obtained from one sample or that obtained from several samples. From this information it is desired to obtain certain characteristics of the parent population. The exact values of these characteristics, of course, cannot be found Yet results can be obtained which appear to be good approximations of the characteristics of the parent.

The quantity $\sqrt{\frac{s(r-1)}{r(s-1)}}$ is less than 1; hence the average of the respective second moments about the means of all possible sample averages is less than the second moment about the mean of the parent. It is necessary to adjust the value of the second moment about the mean of any sample taken at random to make it a better approximation for the second moment about the mean of the parent.

It was proved before that the standard deviation of all possible sample means is

$$\sigma_z = \sigma_v \sqrt{\frac{s-r}{r(s-1)}}, \tag{11.4}$$

when the parent was known. Formula (11.4) gives the standard deviation of all possible sample averages in terms of the standard deviation, σ_v, of the parent. If this, σ_v, were known, the standard deviation of the distribution of the sample means would be known. It is the object of this section to find a similar formula for the standard deviation of the sample means when the parent is not known. Since σ_v, the standard deviation of the parent, is not known the best that can be done is to replace it by something that seems to approximate it. From what has been said above, the best value of the standard deviation of the sample is

$$\sigma_{\text{sample}} = \sigma_{\text{parent}} \sqrt{\frac{s(r-1)}{r(s-1)}},$$

therefore

$$\sigma_{\text{parent}} = \sigma_{\text{sample}} \sqrt{\frac{r(s-1)}{s(r-1)}}.$$

Substituting this value of σ_{parent} in (11.4) for the standard deviation of the parent population gives

$$\sigma_z = \sigma_{\text{sample}} \sqrt{\frac{s-r}{s(r-1)}} = \sigma'_v \sqrt{\frac{s-r}{s(r-1)}}, \tag{11.9}$$

where σ'_v is the standard deviation of the particular sample taken at random. Formula (11 9) gives the best estimate of the standard deviation of the distribution made up of sample averages when the parent is unknown.

When $s \to \infty$ formula (11.9) reduces to

$$\sigma_z = \sigma'_v \sqrt{\frac{1}{r-1}} = \frac{\sigma'_v}{\sqrt{r-1}} = \sqrt{\frac{\Sigma \bar{v}^2}{r-1}} / \sqrt{r} \tag{11.10}$$

or

$$\sigma_z = \frac{s_v}{\sqrt{r}},$$

which is the standard deviation of the sample means when the parent population is not known, and is considered to be the most plausible value. The numerator $s_v = \sqrt{\frac{\Sigma \bar{v}^2}{r-1}}$, in the right-hand member of (11 10) is considered to be the best estimate of the standard deviation of the parent population from which the sample came. The standard deviation of a set of r items v is $\sigma'_v = \sqrt{\frac{\Sigma \bar{v}^2}{r}}$; the best estimate of the standard deviation of the parent from which a sample of r items came is $s_v = \sqrt{\frac{\Sigma \bar{v}^2}{r-1}}$. It is not the exact value and may not be the best that can be obtained, yet it is used as an estimate of the true value. Formula (11.9) is not an exact formula because we do not know how near the standard deviation, σ'_v, of 1 sample, although large, is to the square root of the average of all possible second moments about the various means of the samples. We only assume that it is near it. After all, this is a plausible assumption for large samples.

Formula (11.10) gives the standard deviation of the mean of a sample consisting of r variates, when the parent contains an infinite number of items. Formula (11.10) should be used when the parent is unknown.

Formulas (11.6) and (11.10) are about the same for large values of r.

EXAMPLE. Twenty-seven stalks of wheat were measured with the following results: $M_v = 5.76$ cm., $\Sigma \bar{v}^2 = 86.4$ sq. cm. Find the standard

deviation of the set of measurements, the best estimate of the standard deviation of the parent from which this sample was taken, and the best estimate of the standard deviation of the mean. These values are respectively

$$\sigma'_v = \sqrt{\frac{86\ 4}{27}} = 1\ 789 \text{ cm.},$$

$$s_v = \sqrt{\frac{86\ 4}{26}} = 1\ 823 \text{ cm.},$$

$$\sigma_M = \frac{1\ 823}{\sqrt{27}} = .351 \text{ cm.}$$

$$M_v \pm \sigma_M = 5.76 \quad \pm .351 \text{ cm.}$$

PROBLEMS

1. The length of a table is measured 10 times by A and 10 times by B with the following results:

A	B
Mean = 274.1 in.,	Mean = 274.1 in.,
$\sigma =$ 0.15 in.,	$\sigma =$ 0.66 in.

By means of a diagram, show that the results are not consistent.

2. If a random sample of 36 variates is taken from an infinite parent whose standard deviation is known, how many variates should be taken in another sample to make the standard deviation of the mean of the second sample 6/10 of the standard deviation of the mean of the first sample?

3. The mean and standard deviation of lengths of 26 snakes are respectively: $M_v = 24.8$ in., $\sigma_v = 2.3$ in. Find the standard deviation of this mean, or the distribution of all possible means.

4. If the sum of the squares of the deviations of 37 items from the mean of these 37 items is 144 square feet, find the best estimate of the standard deviation of the parent from which the sample came. Find the standard deviation of the mean. Find the standard deviation of the set of 37 items.

THE STANDARD DEVIATION OF A SUM AND DIFFERENCE

Consider the line ABC. Let n measurements be made of AB and m be made of BC. Let the error of an individual measurement of AB from the mean of all measurements of AB be represented by $\bar{v}$ and that of BC be represented by $\bar{w}$.

A B C

After one measurement has been made of AB, then any one of the m measurements can be made of BC; hence the first error of AB can be added to each of the m errors of BC in finding the error of the sum ABC. The second error of AB can also be added to the m errors of BC in finding the error of ABC, and so on for each of the n errors of AB. The errors $\bar{v}$, $\bar{w}$, and $\bar{v} + \bar{w}$ are listed below:

Error For AB	Error For BC	Error for ABC			
$\bar{v}$	$\bar{w}$	$\bar{v}+\bar{w}$			
$\bar{v}_1$	$\bar{w}_1$	$\bar{v}_1 + \bar{w}_1$	$\bar{v}_1 + \bar{w}_2$	$\bar{v}_1 + \bar{w}_3 \ldots$	$\bar{v}_1 + \bar{w}_m$
$\bar{v}_2$	$\bar{w}_2$	$\bar{v}_2 + \bar{w}_1$	$\bar{v}_2 + \bar{w}_2$	$\bar{v}_2 + \bar{w}_3 \ldots$	$\bar{v}_2 + \bar{w}_m$
$\bar{v}_3$	$\bar{w}_3$	$\bar{v}_3 + \bar{w}_1$	$\bar{v}_3 + \bar{w}_2$	$\bar{v}_3 + \bar{w}_3 \ldots$	$\bar{v}_3 + \bar{w}_m$
$\vdots$	$\vdots$	$\ldots\ldots$	$\ldots\ldots$	$\ldots\ldots$	$\ldots\ldots$
$\bar{v}_m$	$\bar{w}_m$	$\bar{v}_m + \bar{w}_1$	$\bar{v}_m + \bar{w}_2$	$\bar{v}_m + \bar{w}_3 \ldots$	$\bar{v}_m + \bar{w}_m$
$\vdots$		$\ldots\ldots$	$\ldots\ldots$	$\ldots\ldots$	$\ldots\ldots$
$\bar{v}_n$		$\bar{v}_n + \bar{w}_1$	$\bar{v}_n + \bar{w}_2$	$\bar{v}_n + \bar{w}_3 \ldots$	$\bar{v}_n + \bar{w}_m$

Square each error for ABC and add by adding the columns separately; this gives

$$\sum_{i=1}^{n} (\bar{v}_i + \bar{w}_1)^2 = \sum_{i=1}^{n} \bar{v}_i^2 + 2\bar{w}_1 \sum_{i=1}^{n} \bar{v}_i + n\bar{w}_1^2, \text{ for the 1st column,}$$

$$\sum_{i=1}^{n} (\bar{v}_i + \bar{w}_2)^2 = \sum_{i=1}^{n} \bar{v}_i^2 + 2\bar{w}_2 \sum_{i=1}^{n} \bar{v}_i + n\bar{w}_2^2, \text{ for the 2d column,}$$

$$\sum_{i=1}^{n} (\bar{v}_i + \bar{w}_3)^2 = \sum_{i=1}^{n} \bar{v}_i^2 + 2\bar{w}_3 \sum_{i=1}^{n} \bar{v}_i + n\bar{w}_3^2, \text{ for the 3d column,}$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$

$$\sum_{i=1}^{n} (\bar{v}_i + \bar{w}_m)^2 = \sum_{i=1}^{n} \bar{v}_i^2 + 2\bar{w}_m \sum_{i=1}^{n} \bar{v}_i + n\bar{w}_m^2, \text{ for the last column}$$

Now sum all columns; this gives the sum of the squares of all errors. The sum of the squares of all possible errors of measurements of ABC is

$$m\sum_{i=1}^{n} \bar{v}_i^2 + n\sum_{j=1}^{m} \bar{w}_j^2 + 2\sum_{j=1}^{m} \bar{w}_j \sum_{i=1}^{n} \bar{v}_i = m\sum_{i=1}^{n} \bar{v}_i^2 + n\sum_{j=1}^{m} \bar{w}_j^2,$$

since the $\Sigma \bar{v}_{\cdot} = \Sigma \bar{w} = 0$. The standard deviation of the sum is

$$(11.11) \qquad \sigma_{AB+BC} = \sigma_{\text{sum}} = \sqrt{\frac{m\Sigma \bar{v}_i^2}{mn} + \frac{n\Sigma \bar{w}_j^2}{mn}} = \sqrt{\sigma_v^2 + \sigma_w^2},$$

which is equal to

$$\sigma_{ABC} = \sqrt{\sigma_{AB}^2 + \sigma_{BC}^2}.$$

If the signs in the errors for the sums are changed to minus when finding the errors for differences, formula (11.11) will give the standard deviation of the difference, hence

$$(11.12) \qquad \sigma_{\text{difference}} = \sqrt{\sigma_v^2 + \sigma_w^2}.$$

ANOTHER DERIVATION OF THE STANDARD DEVIATION OF A SUM

Let v and w be 2 independently measured quantities with standard deviations respectively, σ_v and σ_w. Let the standard deviation of the sum $v + w$ be σ_{v+w}, the errors from the mean of the measurements made of v be $\bar{v}_1, \bar{v}_2, \bar{v}_3, \ldots, \bar{v}_n$, the errors from the mean of measurements of w be $\bar{w}_1, \bar{w}_2, \bar{w}_3, \ldots, \bar{w}_n$, and the errors for the sum be $\bar{v}_i + \bar{w}_j$. Then

$$\sigma_{v+w} = \sqrt{\frac{\Sigma \bar{v}_i^2}{n} + \frac{\Sigma \bar{w}_j^2}{n} + \frac{2\Sigma \bar{v}_i \bar{w}_j}{n}},$$

$$= \sqrt{\sigma_v^2 + \sigma_w^2},$$

since it is assumed that $\bar{v}_i \bar{w}_j$ will be positive as often as it will be negative and that the sum will be zero or very near to zero. This assumption is plausible when there are a large number of measurements.

In general,

$$\sigma_{\text{sum}} = \sqrt{\sigma_v^2 + \sigma_w^2 + \sigma_q^2 + \ldots + \sigma_s^2}.$$

PROBLEMS

1. The segments AB, BC, CD, DE, EF of the line AF were measured with the following results for the means:

$AB = 7.3$ ft. $\pm$ 0.01,*
$BC = 9.7$ ft. $\pm$ 0 02,
$CD = 4.9$ ft. $\pm$ 0.02,
$DE = 1.8$ ft. $\pm$ 0.03,
$EF = 14.1$ ft. $\pm$ 0.01.

* Plus or minus standard deviation of the mean.

Find the length of AF and its standard error.* If 16 measurements were made of DE, find the standard deviation of the individual measurements. (The standard deviation of a set of items and the standard deviation of each item in the set are the same.)

2. To measure a certain distance, a tape is laid down 529 times, each with a standard error of 0.11 inch. What is the standard error of the whole distance measured?

3. The segments AC and CB, where B is on line AC, were measured with the following results for their means:

$$AC = 39.33 \text{ in.} \pm 0.05,$$

$$CB = 27.89 \text{ in.} \pm 0.03.$$

Find the length of AB and its standard error.

SIGNIFICANCE OF THE DIFFERENCE BETWEEN TWO MEANS

Consider a frequency distribution made up of the weights of men, with mean 140 pounds. The probability that a person, taken at random from this group of men, will weigh more than the mean, 140 pounds, by more than 3 standard deviations is very small. The same is true for picking a person whose weight is less than the mean by 3 standard deviations of the distribution.

Assume that one person took at random 50 men and found the mean of the weights of these 50 men to be 141 pounds, and that another person took a sample of 50 and found the mean weight to be 145 pounds. Are these results significant, or are they due to fluctuations in random sampling? Would one expect a difference as large as this difference in the means to happen often or very seldom? The answer to this question is found in the theory of probability. Set up the distribution of the differences of the means for all possible samples. This distribution will have a mean and a standard deviation as all distributions. Since the distribution of all possible sample averages is normal (when the number in the sample is large†), the distribution made up of the differences of the averages will also be normal. About 68 per cent of this distribution will lie within 1 standard deviation of the mean of this distribution. Since we know from the first of this chapter that the means of the samples lie close to the mean of the parent, the differences of any two sample means will be small, and hence most of these differences will be close to zero. The mean of all possible

* The standard error is the standard deviation of the mean.

† Proved in more advanced works in statistics.

differences between the sample means is zero.* Since the differences of the means of the samples are small and close to zero the standard deviation of the distribution made up of the differences of the means will naturally be small. If the difference of 2 sample means were as large as several standard deviations of this distribution made up of the differences of sample averages, then the probability of getting such a large difference by random sampling would be very small. The answer to this question is obtained by finding the size of this probability. We know that the probability of a variate (that is, one chosen at random) being more than 3 standard deviations from the mean is very small. Hence if the difference between the means is more than 3 standard deviations of the distribution, made up of the differences of all possible means, then the probability is small and the results cannot be due to random sampling. Since the mean of the distribution made up of all possible differences of sample means is zero, the test for significance of the difference between 2 means is found by obtaining a t value as was done on page 88; this test for the significance between 2 means is

(11.13) $S =$

$$\text{Significance} = \frac{\text{Actual difference of the means} - \text{zero}}{\text{Standard error of the difference of the means}}.$$

If this is greater than or equal to 2.6 the difference between the mean is significant and we can say that such a large difference was not due to chance, or did not arise from random sampling from the same parent population. The following problem will bring out the meaning of formula (11.13).

Out of 400 head of cattle, 200 are picked. It is found that the average number of pounds of milk for the first group is 39, for the second group it is 36; also that the best estimates of the standard deviations of the parents from which the first and second groups came are respectively 4.5 and 5.4. Is the difference between the means significant?

The standard error of the first mean is $4.5/\sqrt{200}$.

The standard error of the second mean is $5.4/\sqrt{200}$.

The standard error of the difference of the means, by formula (11.12), is

$$\sqrt{(4.5/\sqrt{200})^2 + (5.4/\sqrt{200})^2} = 0.49.$$

* This is proved in more advanced works in statistics.

Significance = $S = 3/.49 = 6+$ standard errors. Since the difference of the 2 means is over 6 standard errors of the difference of the means from zero, the mean of all possible sample means, the 200 cows were not picked by random sampling. The difference is significant.

PROBLEMS

1. An epidemic resulted fatally in 12.2 per cent of the cases, this being an average of the records from 82 cities, with a standard deviation of 1.3 per cent for the individual records. Upon a recurrence of the epidemic a serum was used and the average this time was 11.0 per cent, and the standard deviation of the individual records was 1.2 per cent. Does the evidence support the use of the serum?

2. At a certain experimental station 50 sheep were fed spelt with an average gain of 25 pounds per animal and standard deviation per animal 9.44 pounds. Another 50 were fed barley with an average gain of 33.9 pounds per animal and standard deviation per animal 2.28 pounds. Is the evidence in favor of the barley conclusive?

3. Two orchards A and B have the same number of trees, n, and the standard deviation of the yield of fruit for the individual trees is 0.51 bushel in both orchards. If orchard A yields on the average 0.15 bushel per tree more than orchard B, how large must n be in order that it may be asserted safely that A is the better orchard?

4. A marksman who is due south of a post has hit the post 41 times in 75 shots. The post subtends at the eye an angle of 2 minutes. Due east of the post stands a cow such that the distance between the cow and the center of the post subtends an angle of 8 minutes at the eyes of the marksman. Is the cow in danger of being shot?

5. The following table gives the frequency distribution of the number of rays per inflorescence of the Golden Alexander (*Zizia aurea*), taken from two different localities. Find the means, the standard deviations of these means, and the significance of the means. Discuss the results.

No. of Rays per Cluster	A Frequency	B Frequency
8–9	0	28
10–11	21	51
12–13	39	72
14–15	98	103
16–17	127	109
18–19	111	71
20–21	66	46
22–23	22	12
24–25	6	8
25–26	11	0

6. Counts of bacteria in 1 cc. of milk:

	Average	Best Estimate of σ of the Parents
First lot of 25 samples ..	150,000	18,000
Second lot of 36 samples..	120,000	12,000

Does the second lot show improvement not attributable to chance?

CHAPTER 12

NON-LINEAR REGRESSION

NON-LINEAR CORRELATION

In Chapters 9 and 10 relations between the variables were assumed to be linear This is not always the case. In physics it is shown that the volume of a gas and the pressure applied to the gas are related according to a law similar to

$$V = C/P, \tag{12.1}$$

where V represents the volume, P the pressure applied, and C some constant. Suppose a physicist has made n observations of the volume of a certain gas after different amounts of pressure have been applied. He has n observed values for V and n corresponding values for P. Substituting these values in (12.1) gives the following n observational equations:

$$V_1 = C/P_1,\ V_2 = C/P_2,\ V_3 = C/P_3, \ldots,\ V_n = C/P_n.$$

To find the "best value" of C it is necessary to find the value which will make the sum of the squares of the residual errors a minimum. The squares of the residual errors are as follows:

$$\begin{aligned}
e_1^2 &= (V_1 - C/P_1)^2 = V_1^2 - 2C\cdot V_1/P_1 + C^2/P_1^2 \\
e_2^2 &= (V_2 - C/P_2)^2 = V_2^2 - 2C\cdot V_2/P_2 + C^2/P_2^2 \\
e_3^2 &= (V_3 - C/P_3)^3 = V_3^2 - 2C\cdot V_3/P_3 + C^2/P_3^2 \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
e_n^2 &= (V_n^2 - C/P_n)^2 = V_n^2 - 2C\cdot V_n/P_n + C^2/P_n^2 \\
\hline
\Sigma e^2 &= \Sigma V^2 - 2C\Sigma V/P + C^2\Sigma(1/P^2) \\
&= \left(\Sigma \frac{1}{P^2}\right) C^2 + 2\left(-\Sigma \frac{V}{P}\right) C + \Sigma V^2.
\end{aligned}$$

The sum of the squares of the residual errors is a quadratic

expression in C; by the lemma on page 147 the quadratic expression is a minimum when

$$\Sigma(1/P^2)C - \Sigma V/P = 0, \text{ or when } C = \frac{\Sigma(V/P)}{\Sigma(1/P^2)}.$$

This value for C is the value which makes the sum of the squares of the residual errors a minimum; hence from the definition this value is the "best value" for the quantity C

The standard error of prediction found as in Chapters 9 and 10 is

$$\sigma_e = \sqrt{\frac{\Sigma V^2 - C\Sigma(V/P)}{n}} = \sqrt{\frac{\Sigma V^2 - (\Sigma V/P)^2/\Sigma(1/P^2)}{n}}.$$

If the quantity $\Sigma\overline{V}^2 + (\Sigma V)^2/n$ is substituted for ΣV^2, the standard error of prediction reduces to

$$\sigma_e = \sigma_y \sqrt{1 - \frac{\dfrac{(\Sigma V/P)^2}{\Sigma(1/P^2)} - \dfrac{(\Sigma V)^2}{n}}{\Sigma\overline{V}^2}} = \sigma_v\sqrt{1 - \mathbf{r}^2},$$

where

$$\mathbf{r} = \sqrt{\frac{\dfrac{(\Sigma V/P)^2}{\Sigma(1/P^2)} - \dfrac{(\Sigma V)^2}{n}}{\Sigma\overline{V}^2}}.$$

The quantity $\mathbf{r}$ is the non-linear correlation coefficient for the particular relation (12.1). In many books it is called the correlation ratio. The relation between σ_e, σ_v and $\mathbf{r}$ may be written as follows:

$$\sigma_e^2 = \sigma_v^2\,(1 - \mathbf{r}^2) \quad \text{or} \quad \mathbf{r}^2 = \frac{\sigma_v^2 - \sigma_e^2}{\sigma_v^2} = 1 - \frac{\sigma_e^2}{\sigma_v^2}.$$

Let y and x be connected by the relation

$$y = a + bx + cx^2, \tag{12.2}$$

and let the n observations for x and the n corresponding observed values for y be substituted in this equation. This gives rise to n observational equations which, according to the least squares method, lead to the following normal equations:

$$\begin{aligned} na + b\Sigma x + c\Sigma x^2 &= \Sigma y \\ \Sigma x\cdot a + b\Sigma x^2 + c\Sigma x^3 &= \Sigma xy \\ \Sigma x^2\cdot a + b\Sigma x^3 + c\Sigma x^4 &= \Sigma x^2 y \end{aligned} \tag{12.3}$$

from which values of a, b, and c can be found If relation (12.2) is written in terms of the deviations from the means the normal equations will be easier to solve.

The standard error of prediction is

$$\sigma_e = \sqrt{\frac{\Sigma y^2 - a\Sigma y - b\Sigma xy - c\Sigma x^2 y}{n}}. \tag{12.4}$$

If $\Sigma \bar{y}^2 + nM_y^2$ is substituted in (12.4) for Σy^2, the standard error of prediction becomes

$$\begin{aligned} \sigma_e &= \sigma_y \sqrt{1 - \frac{a\Sigma y + b\Sigma xy + c\Sigma x^2 y - nM_y^2}{\Sigma \bar{y}^2}} \\ &= \sigma_y \sqrt{1 - \mathbf{r}^2}, \end{aligned} \tag{12.5}$$

where $\mathbf{r}$ is the non-linear correlation coefficient between x and y.

If the predicting equation had been written in terms of the deviations from the means the standard error of prediction could be written as

$$\sigma_e = \sigma_y \sqrt{1 - \frac{b\Sigma \bar{x}\bar{y} + c\Sigma \bar{x}^2 \bar{y}}{\Sigma \bar{y}^2}} = \sigma_y \cdot \sqrt{1 - \mathbf{r}^2} \tag{12.6}$$

where

$$\mathbf{r} = \sqrt{\frac{b\Sigma \bar{x}\bar{y} + c\Sigma \bar{x}^2 \bar{y}}{\Sigma \bar{y}^2}};$$

as before we have

$$\mathbf{r} = \frac{\sigma_y^2 - \sigma_e^2}{\sigma_y^2} = 1 - \frac{\sigma_e^2}{\sigma_y^2}.$$

The quantity $\mathbf{r}$ satisfies the inequalities $-1 \leqq \mathbf{r} \leqq 1$.

PROBLEMS

1. Write out the value of b in (12.3) in terms of the fundamental summations.

2. Set up the normal equations for finding the constants a and b if the predicting equation is $y = a + b\,x^2$.

3. Assume that x and y are connected by the relation $y = a + bx + cx^2$; find the values of a, b, and c which will make the sum of the squares of the residual error a minimum for the following data:

x: 1.3, 2.0, 2.5, 3.0, 3.4, 3.5, 4.0, 5.0, 5.5, 6.0, 7.0, 7.5, 8.0,
y: 2.6, 5.1, 7.3, 9.9, 12.5, 13.3, 16.8, 26.2, 31.3, 36.9, 50.4, 56.6, 73.4,
x: 8.2, 9.0, 9.3, 9.8, 10.
y: 65.4, 81.4, 87.7, 97.3, 101.

Plot the predicting line and the lines parallel to it at a distance of 1 standard error of prediction from it. What percentage of the observed values for y falls in this band? Find the value of y when x is equal to 5.2, also when $x = 8.7$. Find the value of the non-linear correlation coefficient.

4. By means of the lemma on page 147 prove that the normal equations (12.3) arise when the relation between y and x is relation (12.2).

5. The following data give the relation between thread angle in worm gear and efficiency. Plot the data, and find the equation of the curve which seems to fit the data best. Find this equation. Let y represent efficiency.

Angle Thread	Efficiency	Angle Thread	Efficiency
1°	14%	25°	82%
5	52	30	84
7	60	35	86
10	78	40	87
15	75	45	88
20	79	50	87

6. Find the value of **r** which arises when the equation in problem 2 is used for data in problem 5.

CORRELATION RATIO

The non-linear correlation coefficient, as defined above, is obtained after a predicting equation has been assumed; that is, it is a function of the predicting equation. Different predicting equations lead to different values of **r**. When a non-linear correlation coefficient is desired without making any assumption concerning the type of curve which appears to fit the data, it is customary to obtain a coefficient which arises from the average of the squares of the deviations of the y's from the respective means of the y arrays. The quantity σ_e^2, as was shown above, is equal to the average of the squares of the deviations of the y's from the regression line. Let s_y^2 represent the average of the squares of the deviations of the y values from the respective means of the y arrays. From each value of y in the first y array subtract the mean of the first y array; from each y value in the second y array subtract the mean of the second y array; etc. The average of the squares of these deviations from the respective means of the y arrays is equal to s_y^2. The quantity s_y is a standard error of pre-

diction when the predicted values for y are the respective means of the y arrays.* According to this, s_y is†

$$(12.7)\quad s_y = \sqrt{\frac{\sum_{i=1}^{f_1}\left(y_i - \frac{S_1}{f_{y_1}}\right)^2 + \sum_{i=1}^{f_2}\left(y_i - \frac{S_2}{f_{y_2}}\right)^2 + \ldots + \sum_{i=1}^{f_k}\left(y_i - \frac{S_k}{f_{y_k}}\right)^2}{n}}$$

where S_i is the sum of the y's in the ith y array; S_i/f_{y_i} is the mean of the ith y array; f_{y_i} is the number or frequency of the y's in the ith y array; k is the number of y arrays in the distribution; and n is the number in all arrays. This can be written as follows:

(12.8)

$$s_y = \sqrt{\frac{\sum_{i=1}^{f_1} y_i^2 - f_1\left(\frac{S_1}{f_{y_1}}\right)^2 + \sum_{i=1}^{f_2} y_i^2 - f_2\left(\frac{S_2}{f_{y_2}}\right)^2 + \ldots + \sum_{i=1}^{f_k} y_i^2 - f_k\left(\frac{S_k}{f_{y_k}}\right)^2}{n}}$$

$$= \sqrt{\frac{\sum_1^m y^2}{n} - \frac{1}{n}\sum\left(\frac{S_i^2}{f_{y_i}}\right)}.$$

Add and subtract M_y^2 under the radical; this gives

$$s_y = \sqrt{\frac{\Sigma y^2}{n} - M_y^2 - \left[\frac{1}{n}\sum\left(\frac{S_i^2}{f_{y_i}}\right) - M_y^2\right]}$$

$$= \sqrt{\sigma_y^2 - \left[\frac{1}{n}\sum\left(\frac{S_i^2}{f_y}\right) - M_y^2\right]}$$

$$s_y = \sigma_y \cdot \sqrt{1 - \frac{\left[\frac{1}{n}\sum\left(\frac{S_i^2}{f_y}\right) - M_y^2\right]}{\sigma_y^2}},$$

or

$$(12.9)\quad s_y = \sigma_y \cdot \sqrt{1 - \eta^2}, \quad \text{or} \quad \eta_{yx}^2 = 1 - \frac{s_y^2}{\sigma_y^2},$$

where

$$(12.10)\quad \eta_{yx} = \sqrt{\frac{\left[\frac{1}{n}\sum\left(\frac{S_i^2}{f_y}\right) - M_y^2\right]}{\sigma_y^2}}.$$

The quantity η_{yx} is called the correlation ratio and is used as a

* The y array on x is composed of the y values for certain x or a certain x class.

† $f_i = f_{yi}$.

measure of non-linear correlation. The advantage of η_{yx} over **r** is that it does not depend upon any predicting equation. On examining the quantity η_{yx}, it is found to be the standard deviation of the means of the y arrays divided by the standard deviation of the y's for the entire distribution; it can be found from a correlation surface table similar to that on page 178

In a similar way it can be shown that

$$(12.11) \qquad \eta_{xy} = \sqrt{\frac{\left[\frac{1}{n}\sum\left(\frac{T_i^2}{f_{x_i}}\right) - M_x^2\right]}{\sigma_x^2}},$$

where T_i is the sum of the x values in the ith x array on y; T_i/f_{x_i} is the mean of the ith x array; M_x is the mean of the x's for the entire distribution.

The following example will show the difference between r, **r**, η_{yx}, and η_{xy}.

Table 12.1 gives data pertaining to observed tree radii at breast height and areas of cross sections at this height. The radius of a tree at breast height is the average of the maximum and minimum radii. Areas of cross sections are obtained by a planimeter. Values in row S in the above table are found by summing the y' values, which are multiplied by the particular x' in finding the sum of the products, $\Sigma x'y'$. The first value in row S, -16, is found by multiplying $y' = -4$ by its frequency 4. The third number, -41, in this row is found by summing the values of y' which are multiplied by $x' = -2$ in finding the sum of the products $\Sigma x'y'$. There are 3 values of $x' = -2$ and $y' = -4$, 9 values of $x' = -2$ and $y' = -3$, and 1 value of $x' = -2$ and $y' = -2$; hence the sum of these y' values which are to be multiplied by $x = -2$ is $3(-4) + 9(-3) + 1(-2) = -41$; and so on for the other values in row S. The sum of the quantities in the row designated by $S \cdot x'$ is the sum of the product terms, or $\Sigma x'y'$.

Values in column T are obtained in a similar way. The second value, 27, in this column is found by finding the sum of the x' values which are multiplied by $y' = 2$ in finding the sum of the product terms. There are 3 values of $y' = 2$ and $x = 4$ and 5 values of $y' = 2$ and $x' = 3$; hence the sum of these x' values which are multiplied by $y = 2$ in finding the sum of the product terms is $3(4) + 5(3) = 27$. The sum of the values in the column represented by Ty' is the sum of the product terms or $\Sigma x'y'$. The sum of the Ty' column should equal the sum of the Sx' row.

The rows S^2 and S^2/f are used in finding η_{yx}. The quantities η_{yx} and η_{xy} are not always equal. The sum of the values in the row designated

TABLE 12.1

CORRELATION SURFACE TABLE FOR THE RADII OF TREES AND THE AREAS OF THE CROSS SECTIONS AT BREAST HEIGHT

x radius

y = area of cross section

y \ x	2 85–	3 15–	3 45–	3 75–	4 05–	4 35–	4 65–	4 95–	5 25–	$f_{y'}$	y'	fy'	fy'^2	T	Ty'	T^2	$T^2/f_{y'}$
94.5 – 104 5									8	8	3	24	72	32	96	1,024	128 000
84.5 – 94.5								5	3	8	2	16	32	27	54	729	91 125
74 5 – 84 5							12	10	1	23	1	23	23	58	58	3,364	146 260
64 5 – 74 5						18	6	2		26	0	0	0	36	0	1,296	49 846
54.5 – 64.5					23	17	4			44	−1	−44	44	25	−25	625	14 205
44.5 – 54 5			1	15	10	2				28	−2	−56	112	−15	30	225	8 036
34 5 – 44 5		6	9	5	2					22	−3	−66	198	−41	123	1,681	76 409
24.5 – 34 5	4	3	3	1						11	−4	−44	176	−32	128	1,024	93 091
f_x	4	9	13	21	35	37	22	17	12	170		−147	657		464		606 972
x'	−4	−3	−2	−1	0	1	2	3	4								
$f \cdot x'$	−16	−27	−26	−21	0	37	44	51	48	90							
$f \cdot x'^2$	64	81	52	21	0	37	88	153	192	688							
S	−16	−30	−41	−49	−49	−21	8	20	31								
$S \cdot x'$	64	90	82	49	0	−21	16	60	124	461							
S^2	256	900	1,681	2,401	2,401	441	64	400	961								
S^2/f_x	64	100	129 308	114 333	68.600	11 919	2.909	23 529	80 083	594 681							
x'^2y'	−256	−270	−164	−49	0	−21	32	180	496	−52							
fx'^3	−256	−243	−104	−21	0	37	176	459	768	816							
fx'^4	1,024	729	208	21	0	37	352	1,377	3,072	6,820							

by S^2/f_x is part of the first term in formula (12.10). The sum of column T^2/f_y gives a part of the first term in formula (12.11). According to these formulas*

$$\eta^2_{yx} = \left[\frac{594\ 681}{170} - (0.8647)^2\right]/3.117 = 0.8824; \ \eta_{yx} = 0.9394,\dagger$$

and

$$\eta^2_{xy} = \left[\frac{606\ 972}{170} - 0.2803\right]/3.767 = 0.8734; \ \eta_{xy} = 0.9346\cdot\dagger$$

By using data on page 229 and the predicting equation $y = ax^2$, where a is obtained by the least squares method, the non-linear correlation coefficient r is

$$\mathrm{r} = \sqrt{\frac{a\Sigma x^2 y - nMy^2}{\Sigma y^2 - \Sigma y \cdot M_y}},$$

where according to the least squares method

$$a = \frac{\Sigma x^2 y}{\Sigma x^4} = 3.1556.$$

The quantity $\Sigma x^2 y$ is found from the following equation:

$$\text{(12.12)} \qquad \Sigma x^2 y = \Sigma(wx' + h)^2\,(qy' + k) = \Sigma w^2 q x'^2 y' + 2wqh\Sigma x'y' + h^2 q \Sigma y' + kw^2 \Sigma x'^2 + 2hkw\Sigma x' + nh^2 k,$$

where w is the width of the class interval for x, q the width of the class interval for y, h the provisional mean which was subtracted from the mid-points of the x classes, and k the provisional mean which was subtracted from the mid-points of the y classes. From the above table this sum is

$$\begin{aligned}\Sigma x^2 y &= 0.09(10)(-52) + 2(0.3)(10)(4.2)(464) + 10(17.64)(-147) \\ &\quad + 0.09(69.5)(688) + 2(4.2)(0.3)(69.5)(90) \\ &\quad + 170(17.64)(69.5) \\ &= 214{,}197.84.\end{aligned}$$

The quantity Σx^4 is found by the formula

$$\text{(12.13)} \qquad \Sigma x^4 = \Sigma(wx' + h)^4 = w^4 \Sigma x'^4 + 4hw^3 \Sigma x'^3 + 6h^2 w^2 \Sigma x'^2 + 4h^3 w \Sigma x' + nh^4.$$

* In Table 12.1 frequencies of y-arrays are f_x.

† Formulas (12.10) and (12.11) hold for deviations from provisional means, as can be seen by examining formula (12.7).

From the sums of the last two rows in the table and the other computed values this sum according to (12.13) is

$$\Sigma x^4 = 67{,}879.39;$$

hence $a = 3.1556$. The non-linear correlation coefficient $\mathbf{r}$ is

$$\mathbf{r} = \sqrt{.8{,}757} = 0.9358.$$

The values of $\mathbf{r}$ and η_{yx} are nearly the same because the means of the y arrays lie close to the predicting line $y = 3.1556x^2$.

The above example shows that it is much easier to obtain η_{yx} than $\mathbf{r}$.

If the predicting equation is a straight line the linear correlation coefficient is

$$r_{xy} = 0.9295^{+}.$$

Linear regression should not be used here for the data fit a curved line much better than a straight line. The example shows the meaning of r_{yx}, $\mathbf{r}_{yx}$, η_{yx}, and η_{xy}.

The correlation ratio should be used for non-linear correlation, for it gives a measure of dependence based upon the means of the y arrays as the predicted values.

1. For the following data find the linear correlation coefficient and the two correlation ratios.

No. of bracts per inflorescence of *Daucus carota* (wild carrot) for first branch terminal inflorescence

No. of bracts per inflorescence for stem terminals		7	8	9	10	11	12	13	
	13							1	
	12						2		
	11				10	4			
	10	0	1	10	13	7			
	9	1	12	23	14	6			
	8	3	9	11					
	7	2							

2. Rats were fed on a diet containing fixed amounts of vitamin G with the following data pertaining to gains in grams for a certain period.

Amount of Vitamin G	Gain	Amount of Vitamin G	Gain
.5	+ 2	1	14
.5	− 3	1	20
.5	+ 8	1	8
.5	+15	1	17
.5	+ 3	1	26
.5	5	1	18
.5	10	1	16
.5	6		

Amount of Vitamin G	Gain	Amount of Vitamin G	Gain
1.5	20	2	31
1.5	28	2	35
1.5	33	2	36
1.5	+30	2	28
1.5	25	2	29
1.5	30	2	38
		2	36

Find the correlation ratio between amount of vitamin G and gain in weight.

CHAPTER 13

THE ANALYSIS OF TIME SERIES

SECULAR TREND

Time series arise when one considers prices, yields, rates, etc., for a period of time intervals. During each time interval there is a price of a certain commodity, a yield of a certain crop, a rate of exchange of money for a certain country, etc. These values, respectively, form time series. The object of this chapter is to point out some of the ways of analyzing times series. Three characteristics of such series, namely, secular trend, seasonal variations, and cycles, will be introduced together with methods for obtaining them.

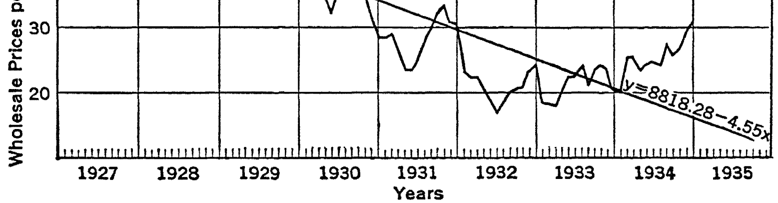


FIG. 13.1.—Average monthly wholesale prices of butter and the straight line trend for yearly averages.

Consider prices of butter for each month from 1928 to 1934 as given in Table 13.1. These prices, plotted in Fig. 13.1, show a downward trend for this period of 7 years including depression years. This trend appears to be linear for most of the data; hence

a straight line was fitted by the method of least squares to the yearly averages given in Table 13.1. This straight line is

$$y = 8{,}818.28 - 4.55x, \tag{13.1}$$

where x represents years and y represents expected yearly averages. This line, shown in Fig. 13 1, indicates the general downward trend of these prices of butter. Equation (13.1) will be used as the trend line for this time series.

A second-degree curve might have been used instead of a straight line, for there is an upward trend in prices of butter from 1932 on. A smooth curve may be drawn in freehand which with care will give a very good trend curve.

The slope of the trend line (13.1) is -4.55, meaning that during 1 year the price of butter dropped 4.55 cents. The drop per month is 1/12 of this, or 0.38 cent.

TABLE 13.1

AVERAGE WHOLESALE PRICES PER POUND OF BUTTER, 92-SCORE CREAMERY AT CHICAGO*

Year	Jan.	Feb	Mar	Apr	May	Jun	July	Aug	Sep.	Oct	Nov	Dec	Ave.
1928	48 76	46 62	49 44	45 49	44 93	44.13	44 93	46 94	48 75	47 79	50.57	50 46	47 40
1929	47 94	49 89	48 45	45.35	43 54	43 54	42 42	43 45	46 22	45 56	42 70	41 10	45.01
1930	36 63	35 70	37 27	38 53	34 85	32.93	35 31	38 92	39 77	39.98	36 09	32 18	36.51
1931	28 50	28.40	28 88	26 10	23 70	23 33	24 95	28 12	32.50	33.76	30 93	30.55	28 31
1932	23 59	22 46	22 61	20 08	18 84	16 99	18 18	20 31	20 76	20.72	23 30	24 11	21.00
1933	18 85	18 65	18 17	20 66	22 54	22 84	24 53	21 31	23 60	24 04	23 60	20 08	21.66
1934	19 84	25 35	25 35	23 66	24 49	24 88	24 49	27 38	25 78	26 93	29 36	30 95	25.71
Ave.	32 16	32 44	32 88	31 41	30 41	29.81	30 69	32 35	33 91	34 11	33 79	32.78	

*Center at middle of month.

SEASONAL VARIATIONS

The average of the January prices is 32.16; the average of the February prices is 32.44. These averages, listed in Table 13.2, contain the effect of the trend. To find the seasonal variations of prices it is necessary to remove the effect of trend from these averages for the monthly prices. To remove trend from the prices it is necessary to add 0.38 cent to the average February prices (since the trend is downward) and 0.76 cent to the average March prices, etc. These averages with the effect of trend removed are given in the second column of Table 13.2. These values are now written as percentages of the average of the values in column 3;

TABLE 13.2

CONSTRUCTION OF INDEXES OF SEASONAL VARIATIONS FROM ARITHMETIC AVERAGES OF ACTUAL PRICES

Month	Average Monthly Prices per Pound	Average Corrected for Secular Trend	Indexes of Seasonal Variations
Jan.	32 16	32 16	93 71
Feb	32 44	32 82	95 63
Mar.	32.88	33 64	98.02
Apr.	31.41	32.55	94.84
May	30.41	31 93	93.04
June	29.81	31 71	92.40
Jul.	30 69	32 97	96.07
Aug.	32 35	35 01	102.10
Sep.	33.91	36 95	107.66
Oct.	34.11	37 53	109.35
Nov.	33.79	35.59	109.53
Dec.	32.78	36.96	107.69
Ave.		34.32	100.00

these are called the indexes of seasonal variation and indicate how the seasons of the years affect prices, or how prices fluctuate for

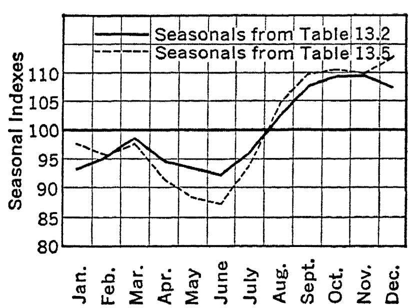


FIG. 13.2.—Seasonal indexes pertaining to prices of butter based on averages of actual prices and chain relatives.

the different periods of the year. These indexes show a decrease from March to June, an increase from June to November, and a decrease from November to January. These indexes of seasonal variation are plotted in Fig. 13.2.

PROBLEMS

1. The following table contains the average price in dollars per 100 pounds of milk received by producers in the United States 1930–1934.

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1930	2.53	2.44	2.38	2 35	2.28	2.22	2.15	2.18	2.25	2.30	2.31	2.20
1931	2.04	1.96	1.92	1 85	1.73	1.66	1 62	1.64	1.70	1.72	1.73	1.67
1932	1.56	1.49	1.43	1.39	1.29	1.17	1.20	1.21	1.25	1 28	1.26	1.26
1933	1.25	1.16	1.10	1.08	1.14	1.21	1.33	1.39	1.47	1 51	1.51	1.49
1934	1.44	1.48	1.50	1.46	1.45	1.47	1.50	1.52	1.57	1.60	1.65	1.69

Find the linear trend line for the yearly averages. Plot this with the data.

2. Find the seasonal variations and plot them.

LINK RELATIVE METHOD OF FINDING SEASONALS

Another method of finding indexes of seasonal variations is by using link relatives. A link relative of prices is the ratio of the price for 1 month to the price of the preceding month. The price of each month must be expressed as a percentage of the price of the preceding month; that is, the January price is expressed as a percentage of the December price, the February price as a percentage of the January price, etc. Table 13.3 contains the link relatives of prices given in Table 13.1. The medians of the link relatives of prices for the various months are recorded in the first column in Table 13.5. These median link relatives measure monthly fluctuations in terms of a shifting base, each monthly price being the base for the next monthly price.

Chain relatives are constructed for the purpose of giving one base monthly price. The median for the link relatives of prices for Januaries will be considered 100. The February median is 97.46 and the February chain relative is also 97.46; the March chain relative is 100.67×97.46, or the median for the March link relatives times the February chain relative. The April chain relative is equal to the April median times the March chain relative, or $93.33 \times 98.11 = 91.57$.* These chain relatives are listed in column 2 of Table 13.5. The December chain relative times the January median should give 100 if no trend were present. The December chain relative times the January median gives

$$103.29 \times 89.12^{*} = 92.05 \text{ instead of } 100,$$

* Divided by 100.

TABLE 13.3—LINK RELATIVES OF PRICES OF BUTTER GIVEN IN TABLE 13.1

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1928	103.13	95.61	106.05−	92.01	98.77	98.21	101.81	104.47	103.86	98.03	105.82	99.78
1929	95.01	104.07	97.11	93.60	96.01	100.00	97.43	102.43	106.38	98.57	93.72	96.25+
1930	98.12	97.46	104.40	103.38	90.45−	94.49	107.23	110.22	102.18	100.53	90.27	89.17
1931	88.56	99.65−	101.69	90.72	90.80	98.44	106.94	112.71	115.58	103.88	91.62	98.77
1932	77.22	95.21	100.67	88.81	93.82	90.18	107.00	111.72	102.23	99.81	112.45+	103.48
1933	82.33	92.95	97.43	113.70	109.10	101.33	107.36	86.87	110.75−	101.86	98.17	85.08
1934	98.80	127.77	100.00	93.33	103.51	101.59	98.43	111.80	101.58	104.46	109.02	105.42
Ave.	90.60	101.96	101.05	96.51	97.49	97.75−	103.74	105.75−	106.08	101.02	100.15+	96.85
Median	89.12	97.46	100.67	93.33	96.01	98.44	106.94	110.22	103.86	100.53	98.17	98.77

TABLE 13.4—FREQUENCY TABLE OF LINK RELATIVES

Classes	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
125–129	..	1	..	..	.	..	..	..	..	..	..	..
120–124	..	..	.	..		..	.	..	..	..	..	..
115–119	..	..	..	..	..	..	..	.	1	..	..	..
110–114	1	..	..	1	..	..	..	4	1	..	1	..
105–109	..	..	1	..	1	..	4	..	1	..	2	1
100–104	..	2	4	1	1	3	1	2	4	5	1	2
95–99	2	3	2	..	2	2	2	..	..	2	..	2
90–94	..	1	..	4	3	2	..	.	..	..	3	..
85–89	2	..	..	1	.	..	..	1	..	..	..	2
80–84	1	..	..	..	..	..	..	..	..	..	..	..
75–79	1	..	..	..	..	..	..		.	.	..	..

TABLE 13.5

INDEXES OF SEASONAL VARIATIONS BY THE METHOD OF LINK RELATIVES

Month	Median Link Relatives	Chain Relatives	Corrected Chain Relatives	Indexes of Seasonal Variations
Jan.	89 12	100	100	97.85
Feb.	97 46	97 46	97 46	95 66
Mar.	100 67	98.11	99 48	97.34
Apr.	93 33	91 57	93 51	91.50
May	96 01	87 92	90 41	88 46
June	98 44	86 55$^-$	89 61	87.68
July	106 94	92.56	96 51	94.43
Aug.	110.22	102 02	107 11	104 80
Sep.	103 86	105.96	112 03	109.62
Oct.	100 53	106 52	113 41	110.97
Nov.	98.17	104 58	112 10	109.69
Dec.	98 77	103.29	114 48	112.02
Jan.		92 05	Ave. 102.20	Ave. 100.00

which shows that trend perhaps has entered into the data and caused this product to differ from 100. During this 12-month period 100 has been reduced to 92.05. It is usually assumed that the rate of increase or decrease is constant for each month. Hence

$$92.05 = 100(1 + r)^{12},$$

or

$$\frac{\log .9205}{12} = \log (1 + r),$$

from which

$$r = -0.0069.$$

To adjust the chain relatives because of the effect of trend, the February chain relative is divided by $1 + r = 0.9931$, the March chain relative by $(1 + r)^2$, the April chain relative by $(1 + r)^3$, etc. These adjusted chain relatives are given in column 4 of Table 13.5. Column 5 of the above table contains the corrected

chain relatives with base equal to the average of the chain relatives in column 4; these are the indexes of seasonal variations and indicate the relative fluctuations in prices for the seasons. These are shown in Fig. 13.2

THE MOVING AVERAGE METHOD OF FINDING SEASONALS

If prices of butter for 12 months, December through November, are averaged, a value is obtained which is centered at June 1; if prices for 12 months, January through December, are averaged, a value is obtained, centered at July 1. Successive 12-month averages are obtained by dropping the prices for the first month, adding a new month and then finding the average. These averages, called moving averages, give a good idea of the trend of prices for the period under consideration.

Table 13.6 gives the sums of prices for the different 12-month periods. Table 13.7 records moving averages centered at the first of the months. Table 13.8 contains the moving averages centered at the fifteenth of the months; these were obtained by averaging two consecutive moving averages in Table 13.7.

TABLE 13.6

SUMS OF PRICES OF BUTTER FOR VARIOUS 12-MONTH PERIODS

Month	1928	1929	1930	1931	1932	1933	1934
Dec.	570 22						
Jan.*	568.81	540.16	439.64	341 20	253.43	261.35	
Feb.	567 99	528.85	431 51	336 29	249 69	261.34	
Mar.	571.26	514.66	424 21	330.35	245.88	268.04	
Apr.	570.27	507.84	415 82	324 08	241 44	275.22	
May	570.13	499.15	403 39	318 06	242 02	278.22	
June	568.74	490.46	392.24	313.20	245.72	280.17	
July	568 15	479.84	382.64	306 86	251 57	282.21	
Aug.	565 64	472.73	372 28	300 09	257 92	282.17	
Sep.	562.15	468.20	361.48	292.28	258.92	288.24	
Oct.	559.62	461.75	354.21	280.54	261.76	290.42	
Nov.	557.39	456.17	347.99	267.50	265.08	293.31	
Dec.	549.52	449.56	342.83	359 87	265.38	299.07	

*Sum for 1928, from January to December, inclusive. Price for December, 1927, is 51.87.

The first figure in column 2 of Table 13.6 is the sum of the prices from December, 1927. through November, 1928; the next figure in

TABLE 13.7

MOVING 12-MONTH AVERAGES OF PRICES, CENTERED AT FIRST OF MONTH

Month	1928	1929	1930	1931	1932	1933	1934
Jan.	. .....	47.35	39 99	31.89	25 57	20.96	23.52
Feb.	......	47.14	39 39	31.02	25 01	21.49	23.51
Mar.	...	46.85	39.02	30.12	24.36	21.58	24.02
Apr.	......	46.64	38 48	29.52	23.38	21.81	24 20
May	.....	46.45	38 01	29.00	22.29	22.09	24 44
June	47.52	45.79	37 46	28.57	21.66	22.12	24.92
July	47.40	45.01	36 64	28 43	21.12	21.78	
Aug.	47 33	44.07	35.96	28 02	20.81	21.78	
Sept.	47.61	42.89	35.35	27.52	20.49	22.34	
Oct.	47 52	42 32	34 65	27.01	20.12	22.94	
Nov.	47 51	41.60	33.62	26.51	20 17	23.19	
Dec.	47.40	40.87	32.69	26.10	20 48	23.35	

TABLE 13.8

MOVING 12-MONTH AVERAGES OF PRICES, CENTERED AT FIFTEENTH OF MONTH

Month	1928	1929	1930	1931	1932	1933	1934
Jan.	....	47.245	39.69	31.455	25.29	21.225	23.515
Feb.	....	46.995	39.205	30.57	24.685	21.54	23.815
Mar.	....	46.795	38.75	29.82	23.87	21.685	24.11
Apr.	......	46.545	38.245	29.26	22.835	21.95	24.32
May	.. ...	46.12	37.735	28.785	21.975	22.105	24.73
June	47.46	45.40	37.50	28.50	21.39	21.95	
July	47.365	44.54	36.30	28.225	20.965	21.78	
Aug.	47.47	43.98	35.655	27.77	20.65	22.06	
Sept.	47.565	42.605	35.00	27.265	20.305	22.69	
Oct.	47.515	41.96	34.135	26.76	20.145	23.065	
Nov.	47.455	41.235	33.155	26.305	20.325	23.27	
Dec.	47.375	40.43	32.29	25.835	20.72	23.435	

this column is the sum of the prices from January, 1928, through December, 1928, etc.

The moving averages in Table 13.8 indicate the trend of prices from 1928 to 1934. This trend is downward from 1928 to the last of 1932; from here to 1934 there is an upward trend of prices. These moving averages, plotted in Fig. 13.3, give a better idea of trend than the straight line used in the first of the chapter.

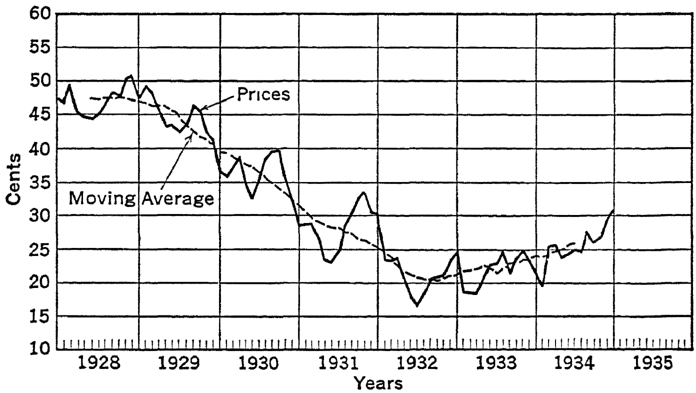


FIG. 13 3—Twelve-month moving average trend line plotted with the original prices of butter.

Table 13.9 contains ratios of the actual monthly prices to the moving averages given in Table 13.8. These values represent seasonal variations with moving-average-trend removed.

The average of the January percentages or ratios in Table 13.9 is 92.59; the average of the February percentages is 95.69, etc. These averages of monthly percentages are listed in Table 13.10, together with their medians.* Indexes of seasonal variations are given in columns 4 and 5 of this table on the basis of the mean and median respectively. There is an increase from January to March, a decrease from March to June, an increase from June to October, and a decrease from October to December.

* The median is the arithmetic average of the middle two.

*

TABLE 13.9

RATIO OF ACTUAL PRICES TO MOVING AVERAGES

Month	1928	1929	1930	1931	1932	1933	1934
Jan.	......	101.47	92.29	90 61	93 28	93.52	84 37
Feb.	....	106.16	91.06	92 90	90 99	86 60	106.43
Mar.	....	103 54	96 18	96 85−	94 72	83 79	105.14
Apr.	.....	97 43	100.75−	89.20	87 94	94 12	97.29
May	....	94 40	92 35+	82.33	85 73	101.97	99 03
June	92.98	95.90	87 81	81.86	79 43	104 05+	
July	94.86	95.24	97 27	88 40	86 72	94 26	
Aug.	98.88	98 79	109 16	101 26	98 35+	96 60	
Sept.	102.49	108.48	113.29	119.20	102 24	104 01	
Oct.	100.58	108.59	117 12	126.16	102 85+	104 23	
Nov.	106 56	103.55−	108 85+	117 58	114 64	101 42	
Dec.	106.51	101 66	99 66	118 25−	116 36	85 68	

TABLE 13.10

INDEXES OF SEASONAL VARIATIONS BY METHOD OF MOVING AVERAGES ON BASIS OF MEAN AND MEDIAN; MOVING-AVERAGE-TREND REMOVED

Month	Average	* Median	Corrected Average	Corrected Median
Jan.	92.59	92.78	93.57	94.72
Feb.	95.69	91.98	96 70	93.90
Mar.	96.70	91.52	97.73	93.43
Apr.	94.46	95.71	95.46	97.71
May	92.64	98.38	93.62	100.44
June	90.34	86.86	91.30	88.68
July	72.79	99.56	93.77	101.64
Aug.	100.51	98.84	101.58	100.91
Sept.	108.29	101.50	109.44	103.62
Oct.	109.92	106.41	111.09	108.63
Nov.	108.77	107.72	109.92	109.97
Dec.	104.69	104.08	105.80	106.26
Ave.	98.95	97.95	100.00	99.99

*The median is the arithmetic average of the middle two.

PROBLEM

1. Find the moving 12-month averages for data on page 236. Set up a table similar to Table 13.10 for these same data. Plot the moving averages.

INDEXES OF SEASONAL VARIATIONS FROM TREND LINE

Table 13.11 contains the values of prices of butter obtained from the trend line $y = 8{,}818.28 - 4.55x$. This trend line was found from the yearly averages which were centered at July 1, the middle of the year. Hence the predicted values are also centered at this date. The predicted value for 1928 is 45.60. From this the predicted value for January 15 is found to be $45.60 + 0.379 \times 5.5 = 47.684$, since the monthly slope of the line is minus 0.379 and there are 5.5 months between July 1 and January 15. The predicted value for February, 1928, is found by subtracting 0.379 from the January predicted price; this is 47.305. The other values in this table are found in a similar way. These values lie on the trend line.

TABLE 13.11

PRICES OBTAINED FROM THE TREND LINE $y = 8{,}818.28 - 4.55x$

Month	1928	1929	1930	1931	1932	1933	1934
Jan.	47 684	43 136	38.588	34.040	29.492	24.944	20.396
Feb.	47.305	42.757	38.209	33.661	29.113	24 565	20.017
Mar.	46.926	42.378	37.830	33 282	28.734	24.185	19.638
Apr.	46.547	41.999	37.451	32.903	28.355	23.807	19.259
May	46.168	41.620	37.072	32.524	27.876	23.428	18.880
June	45.789	41 241	36.693	32.145	27.597	23.049	18.501
July	45.410	40.862	36.314	31.766	27.218	22.670	18.122
Aug.	45.031	40.483	35 935	31.387	26.839	22.291	17.743
Sept.	44.652	40.104	35.556	30.008	26.460	21.912	17.364
Oct.	44.273	39.725	35 177	30.629	26.081	21.533	16.985
Nov.	43.894	39.345	34.798	30.250	25.702	21.154	16.606
Dec.	43.515	38.967	34 419	29.871	25.323	20.775	16.227

Table 13.12 contains ratios of the actual monthly prices to the monthly prices obtained from the trend line. These values are

monthly prices with trend removed, for every price is expressed as a percentage of the trend line value. The ninth column contains the medians for the different months; the last column contains the adjusted medians. These adjusted values are obtained by dividing each median by the average of column 9. The adjusted medians are the seasonal variations, and they indicate how prices fluctuated during the seasons of the year. Averages could be used as well as medians for this purpose. Effect of trend has now been removed from the prices of butter.

TABLE 13.12

Actual Values Divided by Values Obtained from Trend Line

Month	1928	1929	1930	1931	1932	1933	1934	Median	Adjusted median
Jan .	102 27	111 13	94 92	83 73	79 99	79 59	97 25	94 92	94 06
Feb.	98 54	116 67	93 43	84 37	77.16	75 91	126 62	93 43	92 58
Mar.	105 35	114 30	98 52	86 78	78.70	75 11	129 07	98 52	97 63
Apr.	97 72	107 98	102 88	79 33	70.80	86 77	122 85	97 72	96 84
May	97 31	104 61	94 01	72.88	67.36	96.20	129 71	96 20	95.33
June.	96 37	105 58	89 75	72 57	61 56	99 09	134.49	96 37	95 50
July..	98 72	103 82	97 25	78.53	66 79	108 20	135 15	98 72	97.83
Aug.	104 24	107 34	108 29	89 58	75 67	95 60	154 34	104 24	103.30
Sept .	109 18	115 26	111.84	108.30	78 46	107.71	148 50	109 18	108.19
Oct...	107 95	114 67	100 63	110 87	79 45	111 66	158 51	110 89	109.87
Nov..	115 22	108 51	103 71	102 25	90.66	111.58	176 76	108 51	107.53
Dec.	115 95	105 47	93 49	102 28	95 22	96 63	190.70	102 28	101.35

It is now possible to remove seasonal variations from the data by subtracting the seasonal variations from the data after trend has been removed. This is done in Table 13.13. Here the seasonal variations in the third column are those in the last column in Table 13.12. Prices of butter with trend and seasonals removed are found in column headed $a - b$; these are expressed as differences of percentages. Notice that the seasonals are repeated every 12 months as the fourth column shows.

TABLE 13.13

CYCLE FLUCTUATIONS AFTER TREND AND SEASONAL VARIATIONS HAVE BEEN REMOVED FROM PRICES OF BUTTER

Year	Month	Ratio of Actual Prices to Prices from Trend Line, a	Seasonal Variations, b	Deviations, $a - b$	Deviations Squared, $(a - b)^2$	Deviations in Terms of Standard Deviations
1928	Jan.	102 27	94.06	8 21	67.4041	0 37
	Feb.	98 54	92.58	5 96	35.5216	0 28
	Mar.	105 35	97.63	7.72	59.5984	0 35
	Apr.	97 92	96 84	0.88	0.7744	0 04
	May	97 31	95.33	1.98	3.9204	0 09
	June	96 37	95 50	0 87	0.7569	0 04
	July	98 72	97.83	0.89	0.7921	0.04
	Aug.	104 24	103 30	0 94	0.8836	0 04
	Sept.	109 18	108 19	0.99	0.9801	0 04
	Oct.	107 95	109 87	−1.92	3.6864	−0.09
	Nov.	115.22	107.53	7.69	59.1361	0.34
	Dec.	115.95	101.35	4.60	21.1600	0.21
1929	Jan.	111.13	94 06	17.07	291.3849	0.77
	Feb.	116.67	92 58	24 09	580.3281	1.08
	Mar.	114 30	97 63	16 67	277.8889	0.75
	Apr.	107.98	96 84	11.14	124.0996	0.50
	May	104 61	95 33	9 28	86.1184	0.42
	June	105.58	95.50	10 08	101 6064	0.45
	July	103 82	97.83	5 99	35 8801	0 27
	Aug.	107.34	103.30	4.04	16 3216	0.18
	Sept.	115 26	108.19	7 07	49.9849	0.32
	Oct.	114.67	109 87	4 80	23 0400	0.22
	Nov.	108 50	107 53	0.98	0.9604	0.04
	Dec.	105.47	101 35	4.12	16.9744	0.18
1930	Jan.	94.92	94 06	0.86	0.7396	0.22
	Feb.	93.43	92 58	0.85	0.7225	0.04
	Mar.	98 52	97 63	0.89	0.7921	0 04
	Apr.	102.88	96 84	6.04	36 4816	0.27
	May	94 01	95 33	−1.32	1.7424	−0.06
	June	89.75	95 50	−5.75	33 0625	−0 26
	July	97.25	97.83	−0 58	0 3364	−0.03
	Aug.	108 29	103.30	4 99	24.9001	0.22
	Sept.	111.84	108 19	3 65	13.3225	0.16
	Oct.	100 63	109.87	−9.24	85.5625	−0.41
	Nov.	103 71	107 53	−3.82	14 5924	−0.17
	Dec.	93 49	101.35	−7.86	61.7796	−0.35
1931	Jan.	83.73	94.06	−10.33	106.7089	−0.46
	Feb.	84 37	92.58	−3 26	10.6276	−0 15
	Mar.	86.78	97.63	−10.06	101.2036	−0 45
	Apr.	79 33	96 84	−17 51	306.6001	−0 78
	May	72 88	95 33	−22 45	504.0025	−1.01
	June	72 57	95.50	−22 93	525.7849	−1.03

TABLE 13.13—(*Continued*)

CYCLE FLUCTUATIONS AFTER TREND AND SEASONAL VARIATIONS HAVE BEEN REMOVED FROM PRICES OF BUTTER

Year	Month	Ratio of Active Prices to Prices from Trend Line, a	Seasonal Variations, b	Deviations, $a - b$	Deviations Squared, $(a - b)^2$	Deviations in Terms of Standard Deviations
1931	July	78 53	97 83	−19 30	372 4900	−0 86
	Aug.	89 58	103 30	−13 72	188 2384	−0 61
	Sept	108 30	108 19	0 11	0 0121	0 00
	Oct.	110 87	109 87	1 00	1 0000	0 04
	Nov.	102 25	107 53	−5 28	27 8784	−0 24
	Dec.	102 28	101 35	0 93	0 8649	0 04
1932	Jan.	79 99	94 06	−14 07	197 9649	−0 63
	Feb	77 16	92 58	−15 42	237 7764	−0 69
	Mar.	78 70	97 63	−18 93	358 3449	−0 85
	Apr.	70 80	96 84	−26 04	678 0816	−1 17
	May	67 36	95 33	−27 97	782 3209	−1 25
	June	61 56	95 50	−33 94	1,151 9236	−1 52
	July	66 79	97 83	−31 04	963 4816	−1 39
	Aug.	75 67	103 30	−27 63	763 4169	−1 24
	Sept.	78 46	108 19	−29 73	883 8729	−1 33
	Oct.	79 45	109 87	−30 42	925 3764	−1 36
	Nov.	90 66	107 53	−16 87	284 5969	−0 76
	Dec.	95 22	101 35	−6 13	37 5769	−0 27
1933	Jan.	79 59	94 06	−14 47	209 3809	−0 65
	Feb.	75 91	92 58	−16 67	277 8889	−0 75
	Mar.	75 11	97 63	−22 52	507 1504	−1 01
	Apr.	86 77	96 84	−10 07	101 4049	−0 45
	May	96 20	95 33	0 87	0 7569	0 04
	June	99 09	95 50	3 59	12 8881	0 16
	July	108 20	97 83	10 37	107 5369	0 46
	Aug	95 60	103 30	−7 70	59 2900	−0 34
	Sept.	107 71	108 19	−0 48	0 2304	−0 02
	Oct.	111 66	109 87	1 79	3 2041	0 08
	Nov.	111 58	107 53	4 05	16 4025	0 18
	Dec.	96.63	101 35	−4 72	22 3729	−0.21
1934	Jan.	97.25	94 06	3 19	10 1761	0 14
	Feb.	126 62	92 58	34 04	1,158 7216	1 50
	Mar.	129.07	97 63	31 44	988 4736	1 38
	Apr.	122 85	96 84	26 01	676 5201	1 14
	May	129 71	95 33	34 38	1,178 5464	1 51
	June	134 49	95 50	38 99	1,520 2201	1.72
	July	135 15	97 83	37 32	1,392 7824	1 64
	Aug.	154 34	103.30	51 04	2,605 0816	2.25
	Sept.	148 50	108 19	40 31	1,624.8961	1 77
	Oct.	158.51	109 87	57 64	3,322.3696	2.54
	Nov.	176 76	107 53	69.23	4,792 7929	3.05
	Dec.	190 70	101 35	89 35	9,672 7225	3 93

$$\sigma = \sqrt{\frac{41{,}805.1912}{84}} = \sqrt{497.6808} = 22.31,$$

found from the total of column 6.

CYCLE FLUCTUATIONS

Time series are usually influenced by trend, seasonal variations, cycle fluctuations, and residuals which cannot always be measured. After the effects of trend and seasonal variations have been removed from the original data there remain cyclical fluctuations and residuals which should be located and analyzed if possible. One way of determining cycle variation is to eliminate the effect of seasonal variations from trend line values by correcting the ratios of actual prices to the ordinates of the trend line for seasonals. The necessary computations are presented in Table 13.13. Column 5 contains the data with trend and seasonals removed. These values are plotted in Fig. 13.4.

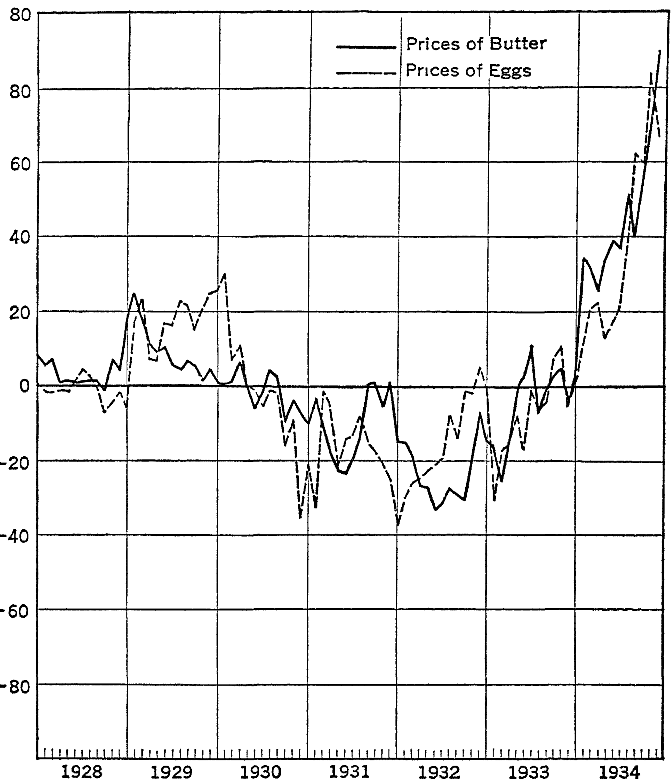


FIG. 13.4.—Wholesale prices of butter and eggs with trends and seasonal variations removed, leaving cyclical and residual variations.

The last column in Table 13 13 gives the size of the deviations from zero in terms of the standard deviations. Some business firms like to plot these values instead of the values $a - b$ There do not appear to be definite cycles for this period. Cycles are difficult to detect in time series when the time under consideration extends over a short period. A cycle chart pertaining to one commodity can be compared with cycle charts of other commodities by plotting on the same axes or by placing them over a glass under which is a bright light.

Figure 13 4 shows the cyclical variations and residual errors pertaining to prices of eggs after trend and seasonals have been removed. Cycles pertaining to butter prices nearly coincide with those pertaining to prices of eggs. The broken lines in this figure follow about the same general course during this period of seven years.

PROBLEMS

Analyze the following time series, securing secular trend, seasonal variations, and cyclic variations.

1. Average prices per 100 pounds paid producers by condensaries for milk testing 3.5 butterfat, f.o.b. factory. Part of Table 416 in *Agricultural Statistics*, 1937.

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1932	1.12	0.99	0.95	0.93	0 86	0.81	0.77	0.80	0.85	0.86	0.86	0.92
1933	0.95	0.84	0.82	0.81	0.93	1.00	1.07	1.10	1.09	1.10	1.08	1.00
1934	0.97	1.10	1.11	1.02	1.06	1.09	1.09	1.21	1.17	1.20	1.32	1.35
1935	1.46	1.57	1.42	1.46	1.23	1.13	1.13	1.18	1.22	1.31	1.49	1.57
1936	1.58	1.62	1.47	1.40	1.29	1.39	1.63	1.74	1.74	1.65	1.64	1.62

2. Butterfat: Average price per pound received by farmers, U.S.A. Part of Table 420 in *Agricultural Statistics*.

Year	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1932	22.8	19.8	19.5	17.8	16.3	14.6	14.4	17.5	17.6	17.8	18.4	21.1
1933	18.9	15.8	15.1	16.5	20.2	19.7	23.0	18.4	19.6	20.1	20.4	18.0
1934	16.1	21.6	23.5	21.0	21.5	22.2	22.1	24.3	24.0	24.3	27.2	28.2
1935	30.5	35.9	31.2	33.8	27.5	23.7	22.3	22.9	24.9	25.9	29.9	33.0
1936	33.5	34.9	31.7	31.2	27.1	27.7	32.6	35.7	35.5	33.5	33.1	33.6

CHAPTER 14

ANALYSIS OF VARIANCE

SUMS OF SQUARES

In Chapter 11 a method was given for finding the significance of 2 arithmetic means. The present chapter presents methods of testing the significance of 2 or more means by the analysis of variances. Consider the following example of wheat yields of 3 varieties, A, B, and C, each of which was planted on 5 plots. Table 14.1 contains the yields for the plots.

TABLE 14.1

Yields of Wheat

	A	B	C	
	36	34	26	
	33	24	27	
	40	39	29	
	31	26	24	
	35	27	34	
Total	175	150	140	465
Mean	35	30	28	31

If there are no differences between the varieties the means of the yields should be about the same; that is, the differences should be small and due only to fluctuations arising in random sampling. To be able to test whether or not there are significant differences between means it is necessary to find standard errors and standard errors for differences of means. These standard errors are found from the analysis of variances, as will be explained.

The sum of the squares of the deviations of the plot yields from the mean of all the yields will be designated by the phrase *total* sum of squares. This *total* sum of squares can be broken up into two sums of squares, namely, the sum of squares of the deviations of the variety means from the mean of all yields, and the sum of the squares of the deviations of the plot yields from their corresponding variety means. The *total* sum of squares can be broken up into the sum of squares *between variety means* and the sum of squares *within varieties.*

Let x_{ij} represent the yield from the jth plot of the ith variety, $\bar{x}$ the mean of all plot yields, $\bar{x}_1$* the mean of the yields of variety A, $\bar{x}_2$ the mean of the yields of variety B, $\bar{x}_3$ the mean of yields of variety C, or $\bar{x}_p$ $(p = 1, 2, 3)$ represents the means of varieties. The relation between these sums of squares can be expressed as

$$(14.1) \qquad \Sigma\,(x_{ij} - \bar{x})^2 = \Sigma\,(\bar{x}_p - \bar{x})^2 + \Sigma\,(x_{ij} - \bar{x}_p)^2.$$

The quantity $\Sigma(x_{ij} - \bar{x})^2$ is the total sum of squares, the quantity $\Sigma(\bar{x}_p - \bar{x})^2$ is the sum of squares *between variety means,* and the quantity $\Sigma(x_{ij} - \bar{x}_p)^2$ is the sum of squares *within varieties.* For the data in Table 14.1 these sums of squares are as follows:

$$\Sigma(x_{ij} - \bar{x})^2 = (36 - 31)^2 + (33 - 31)^2 + \cdots + (24 - 31)^2 + (34 - 31)^2$$
$$= 392.$$

$$\Sigma(\bar{x}_p - \bar{x})^2 = 5(35 - 31)^2 + 5(30 - 31)^2 + 5(28 - 31)^2 = 130.$$

$$\begin{aligned}\Sigma(x_{ij} - \bar{x}_p)^2 = {} & (36 - 35)^2 + (33 - 35)^2 + \cdots + (35 - 35)^2 + (34 - 30)^2 \\ & + (24 - 30)^2 + \cdots + (27 - 30)^2 + (26 - 28)^2 \\ & + (27 - 28)^2 + \cdots + (34 - 28)^2 = 262.\end{aligned}$$

In the first sum in the right member of equation (14.1) the mean $\bar{x}_p = \bar{x}_1$ for items in the first column, $\bar{x}_2$ for items in the second column, and $\bar{x}_3$ for items in the third column; thus each of the quantities $(\bar{x}_1 - \bar{x})^2$, $(\bar{x}_2 - \bar{x})^2$, and $(\bar{x}_3 - \bar{x})^2$ is repeated 5 times. In the last sum of squares in equation (14.1), $\bar{x}_p = \bar{x}_1$, when x_{ij} takes values in the first column; $\bar{x}_p = \bar{x}_2$, when x_{ij} takes values in

* The notation will be changed for this chapter because almost all writers in their work in the analysis of variances use these notations.

the second column; and $\bar{x}_p = \bar{x}_3$, when x_{ij} takes values in the third column Equation (14.1) can be written for the general case as follows:

$$(14.2) \qquad \Sigma(x_{ij} - \bar{x})^2 = k\Sigma(\bar{x}_p - \bar{x})^2 + \Sigma(x_{ij} - \bar{x}_p)^2,$$

where k is the number of values from which $\bar{x}_p$ comes.

Equation (14.1) may be proved to be true from the identity

$$(14.3) \qquad (x_{ij} - \bar{x}) = (\bar{x}_p - \bar{x}) + (x_{ij} - \bar{x}_p).$$

Square both members of this identity and sum for all values of x_{ij}; this gives

$$(14.4) \quad \Sigma(x_{ij} - \bar{x})^2 = \Sigma(\bar{x}_p - \bar{x})^2 + 2\Sigma(\bar{x}_p - \bar{x})(x_{ij} - \bar{x}_p) + \Sigma(x_{ij} - \bar{x}_p)^2.$$

The middle term in the right member of (14.4) is equal to zero; for consider the simple table with three varieties and two replications.

A	*B*	*C*
x_{11}	x_{21}	x_{31}
x_{12}	x_{22}	x_{32}

Consider the middle term in the right member of (14.4) for the above table; this term expanded is

$$\begin{aligned} 2\Sigma(\bar{x}_p - \bar{x})(x_{ij} - \bar{x}) &= 2(\bar{x}_1 - \bar{x})(x_{11} - \bar{x}_1) + 2(\bar{x}_1 - \bar{x})(x_{12} - \bar{x}_1) \\ &+ 2(\bar{x}_2 - \bar{x})(x_{21} - \bar{x}_2) + 2(\bar{x}_2 - \bar{x})(x_{22} - \bar{x}_2) \\ &+ 2(\bar{x}_3 - \bar{x})(x_{31} - \bar{x}_3) + 2(\bar{x}_3 - \bar{x})(x_{32} - \bar{x}) \\ &= 2(\bar{x}_1 - \bar{x})(0) + 2(\bar{x}_2 - \bar{x})(0) + 2(\bar{x}_3 - x)(0) \\ &= 0, \end{aligned}$$

since the sum of the items in any column is equal to the mean of the items times the number of items. Equation (14.4) reduces to equation (14.2) when $k = 2$.

The sum of squares *within varieties* is usually found by subtracting the sum of squares *between means* from the *total* sum of squares. Table 14.2 gives the analysis of variances for the wheat yields.

TABLE 14.2

ANALYSIS OF VARIANCES FOR WHEAT YIELDS

Source of Variation	Sums of Squares	Degrees of Freedom	Variance	Standard Deviation
Total..... .	392	14	28 000	5 292
Between means of varieties...	130	2	65	8.062
Within varieties	262	12	21 833	4 673

The variances in column 4 are found by dividing the sums of squares by the corresponding degrees of freedom. The degrees of freedom for the *total* variance is 1 less than the number of plot yields, or $15 - 1 = 14$; the degrees of freedom for the *between means* variance is 1 less than the number of varieties, or $3 - 1 = 2$; the degrees of freedom for the *within varieties* variance is found by subtracting the last number of degrees of freedom from the former, or by the product $h(k - 1)$, where k is the number of rows in each column and h is the number of columns. Each of the 3 variances, provided the data are homogeneous, are independent estimates of the variance of the assumed normal parent population from which this sample was taken. The *within-varieties* variance is usually taken as the estimate of the variance of the population from which the yields were taken. It is found from the sum of squares which is obtained by subtracting from the *total* sum of squares the sum of the squares *between means*; that is, it is obtained after subtracting out the variations due to different means. The standard deviation, or an estimate of the standard deviation of the parent from which the sample came, is considered to be the square root of the variance *within varieties*; for the wheat yields it is 4.673. The standard deviation of any of the means is equal to

$$\sigma_{\text{mean}} = \frac{\text{Estimated } \sigma \text{ of the parent}}{\sqrt{N}}$$

$$= \frac{\sigma \text{ for } \textit{within varieties}}{\sqrt{N}} = \frac{4.67}{\sqrt{5}} = 2.09,$$

where N is the number from which the mean was found. Means can now be tested as in Chapter 11. Snedecor's table enables one to test for significance between the means by finding a value F where

$$F = \frac{\text{Larger variance}}{\text{Smaller variance}} = \frac{65}{21.833} = 2.98.$$

Is this ratio too large to be attributable to fluctuations due to chance? Snedecor's table, Table III in this book, shows that F should be greater than 3.88 to be significant. Thus there are no significant differences, between the means of the varieties, and as far as this experiment goes one variety is as good as each of the others. The number 3.88 is found in Table III by going down the second column (2 degrees of freedom) to the twelfth row (12 degrees of freedom) for the variance in the denominator of F.

In the analysis of variance all the data are pooled in finding the variance *within varieties*; this gives a much better estimate of the variance of the parent population than using only 5 yields. The variance *within varieties* is used as experimental error because the variations due to different means have been subtracted from the *total* sum of squares. The value of F shows whether there are significant differences between the means, and in the above example its size showed that there were no significant differences between any 2 of the means. If the value of F suggests no significant differences between any 2 means it is useless to test further; F tests significance between variances and suggests significance or no significance between means.

RANDOMIZED BLOCKS

Consider the case where 5 varieties of beans, A, B, C, D, and E are grown on 4 blocks of land, each of which is divided into 5 equal plots. Varieties are assigned to the plots in each block at random; this might be done by drawing tickets from a box, where each variety is represented by a ticket. The first variety drawn is planted in the first plot in the block under consideration, the next variety drawn is planted in the second plot of this block, etc. Each block contains 1 plot of each variety; thus each variety is replicated 4 times in this example. The following chart gives an arrangement of the plots.

Blocks			Plots		
1	*A*	*D*	*E*	*C*	*B*
2	*B*	*C*	*A*	*E*	*D*
3	*A*	*E*	*D*	*B*	*C*
4	*E*	*B*	*C*	*D*	*A*

The following table gives yields in grams per plot for beans arranged for computations.

TABLE 14.3

YIELDS OF 5 VARIETIES OF BEANS

Blocks	Varieties					Totals
	A	*B*	*C*	*D*	*E*	
1	42	36	34	46	32	190
2	38	34	30	39	37	178
3	47	42	37	42	36	204
4	41	40	39	37	35	192
Totals	168	152	140	164	140	764
Means	42	38	35	41	35	38 2

The total sum of squares of the deviations of the yields from the general mean can be broken up into three sums of squares, namely, the sum of squares of the deviations of the means of the varieties from the general mean, the sum of squares of the deviations of the means of the blocks from the general mean, and the sum of squares due to error. The equation which expresses this relation is

$$\Sigma(x_{ij} - \bar{x})^2 = 4\Sigma(\bar{x}_v - \bar{x})^2 + 5\Sigma(\bar{x}_b - \bar{x})^2 + \Sigma(x_{ij} - \bar{x}_v - \bar{x}_b + \bar{x})^2 \tag{14.5}$$

where $\bar{x}_v$ represents the variety means and $\bar{x}_b$ represents the block means. The quantity $\bar{x}_b$ is equal to the mean of the third block when x_{ij} takes values, in the summations, in the third block, etc.

The last summation in the above equation is usually found by subtracting the first two from the total sum of squares. Formula (14.5) can be proved as Formula (14.1). Equation (14.5) may be written as

Total sum of squares = Sum of square *between variety means*
+ Sum of squares *between block means*
+ Sum of squares due to *error*.

The sum of squares $4\Sigma(\bar{x}_v - \bar{x})^2$ can be written as follows:

$$4\sum_{v=1}^{5}(\bar{x}_v - \bar{x})^2 = 4\sum_{v=1}^{5}(\bar{x}_v - 2\bar{x}_v\bar{x} + \bar{x}^2) = 4\left(\sum \bar{x}_v{}^2 - 2\bar{x}\cdot\sum_{1}^{5}\bar{x}_v + 5\bar{x}^2\right)$$

$$= 4(\Sigma\bar{x}_v{}^2 - 2\bar{x}\cdot 5\;\frac{\Sigma\bar{x}_v}{5} + 5\bar{x}^2) = 4(\Sigma\bar{x}_v{}^2 - 10\cdot\bar{x}^2 + 5\bar{x}^2)$$

$$= 4(\Sigma\bar{x}_v{}^2 - 5\cdot\bar{x}^2)$$

$$= 4(\bar{x}_1{}^2 + \bar{x}_2{}^2 + \bar{x}_3{}^2 + \bar{x}_4{}^2 + \bar{x}_5{}^2 - 5\bar{x}^2)$$

$$= 4\left[\left(\frac{\overset{\mathrm{I}}{\Sigma} x}{4}\right)^2 + \left(\frac{\overset{\mathrm{II}}{\Sigma} x}{4}\right)^2 + \left(\frac{\overset{\mathrm{III}}{\Sigma} x}{4}\right)^2 + \left(\frac{\overset{\mathrm{IV}}{\Sigma} x}{4}\right)^2 + \left(\frac{\overset{\mathrm{V}}{\Sigma} x}{4}\right)^2\right]^*$$

$$- 20\bar{x}^2$$

$$= \frac{T_A^2 + T_B^2 + T_C^2 + T_D^2 + T_E^2}{4} - 20\left(\frac{\sum^{n} x_{ij}}{20}\right)^2$$

$$= \frac{\sum_{A}^{E} T_i{}^2}{4} - \frac{(\Sigma x)^2}{20},$$

$$(i = A, B, C, D, E)$$

where T_B is the total of the yields for variety B, etc. The term $(\overset{n}{\Sigma} x)^2/20$ is called the correcting factor or term. The sum of squares *between block means* may be written in a similar manner as

$$5\sum_{b=1}^{4}(\bar{x}_b - \bar{x})^2 = \frac{\sum_{1}^{4} t_j{}^2}{5} - \frac{\left(\sum^{n} x\right)^2}{20},$$

* The values above the summation signs indicate that these sums take in variety A, B, etc.

t_2 represents the totals of the second block. The correcting factor is the same for all the sums of squares.

The total sum of squares is equal to

$$\Sigma(x - \bar{x})^2 = 42^2 + 38^2 + \ldots + 36^2 + 35^2 - \frac{(764)^2}{20} = 359.2,$$

the sum of squares *between means* of blocks is

$$5\Sigma(\bar{x}_b - \bar{x})^2 = \frac{\Sigma t_j^2}{5} - \frac{(\Sigma x)^2}{20} = \frac{190^2 + 178^2 + 204^2 + 192^2}{5} - \frac{(764)^2}{20} = 68.0,$$

the sum of squares between *variety means* is

$$4\Sigma(\bar{x}_v - \bar{x})^2 = \frac{\Sigma T_i^2}{4} - \frac{(\Sigma x)^2}{20} = \frac{168^2 + 152^2 + 140^2 + 164^2 + 140^2}{4} - \frac{(764)^2}{20} = 171.2.$$

The sum of squares due to *error* is found by subtracting the sum of the last 2 sums of squares from the *total* sum of squares.

Table 14.4 gives the analysis of variances for the yields of beans.

TABLE 14.4

ANALYSIS OF VARIANCE

Source of Variation	Sum of Squares	Degrees of Freedom	Variance
Total	359.2	19	18.91
Between means of varieties	171.2	4	42.80
Between means of blocks	68 0	3	22.67
Error	120.0	12	10.00

The error variance is commonly used as the estimate of the variance of the parent population from which the sample of bean

yields was taken at random. By comparing the error variance with the variance *between means* of blocks it is seen that F is

$$F = \frac{22.67}{10.00} = 2.27.$$

From Table III, F should be as large or larger than 3.49 to be significant; that is, there are no significant differences *between the means* of the blocks. This shows that the soil was rather homogeneous. Comparing the error variance with the variance *between means* of varieties, F is

$$F = \frac{42.80}{10.00} = 4.80,$$

which is too large to be the result of random sampling from the same parent: hence, there is a significance between at least two of the variety means. Let us test for the difference between the means of variety A and variety C. The standard error of any of the means of the varieties is

$$\sigma_{\text{mean}} = \sqrt{\frac{\text{Error variance}}{4}} = \sqrt{\frac{10\ 00}{4}} = 1.58$$

and the standard error of the difference of the means is

$$\sigma_{\text{difference of means}} = \sqrt{(1.58)^2 + (1.58)^2} = 1.58\sqrt{2} = 2.23.$$

$$t = \frac{\text{Mean of } A\text{—Mean of } C}{2.23} = 3.14.$$

The values of t at the 5% level and the 1% level are given in Table III. Variety A is better than variety C or E. Other means can be tested in this way. A difference between two means as great as 4.8 grams is significant. Variety D is significantly better than varieties C and E. The 4.8 is found by solving the equation.

$$\frac{\text{Difference of means}}{2.23} = 2.179.$$

The number 2.179 is found in the t column in Table III at the 5% level for 12 degrees of freedom for *error*.

The sums of squares can be found by using a provisional mean. This enables one to reduce the sizes of the items and hence to reduce

the amount of computing. Consider Table 14.3. Subtract 40 from each value. These new data are recorded in Table 14.6.

TABLE 14.6

YIELDS OF 5 VARIETIES OF BEANS LESS 40

Blocks	Varieties					Totals
	A	*B*	*C*	*D*	*E*	
1	+2	−4	−6	6	−8	−10
2	−2	−6	−10	−1	−3	−22
3	7	2	−3	2	−4	4
4	1	0	−1	−3	−5	−8
Totals	8	−8	−20	4	−20	−36

The sum of squares

$$\Sigma(x - \bar{x})^2 = 2^2 + (-2)^2 + \ldots + (-4)^2 + (-5)^2 - (-36)^2/20$$
$$= 359.2, \text{ as found before.}$$

The sum of squares

$$4\Sigma(\bar{x}_v - \bar{x})^2 = \frac{8^2+(-8)^2+(-20)^2+4^2+(-20)^2}{4} - (-36)^2/20$$
$$= 171.2 \text{ as found before.}$$

The sum of square $5\Sigma(\bar{x}_b - \bar{x})^2 = \dfrac{(-10)^2 + (-22)^2 + 4^2 + (-8)^2}{5}$

$$- \frac{(-36)^2}{20} = 68.0, \text{ as found before.}$$

The sum of squares for *error* is found as before.

THE LATIN SQUARE

Consider a field experiment carried out on a piece of land divided into plots which are arranged in rows and columns, where there are the same number of rows as columns and where each

variety or treatment occurs once in each row and once in each column The varieties or treatments are assigned to the plots at random. An example of a 4×4 Latin square is shown in the following diagram, where A, B, C, and D represent treatments.

A	*C*	*B*	*D*
C	*B*	*D*	*A*
D	*A*	*C*	*B*
B	*D*	*A*	*C*

The following example pertaining to the heights of oats grown in four treatments, will help to explain how the Latin square is used to test significance between means of varieties or treatments.

HEIGHTS OF OATS IN CM. FROM PLOTS ARRANGED IN A LATIN SQUARE

	Columns				Total Rows	Total Treatments	Treatment Means
Rows	*A* 18	*C* 19	*B* 11	*D* 15	63	$A = 60$	*A* 15
	C 20	*B* 9	*D* 16	*A* 16	61	$B = 47$	*B* 11 75
	D 17	*A* 14	*C* 17	*B* 13	61	$C = 71$	*C* 17.75
	B 14	*D* 13	*A* 12	*C* 15	54	$D = 61$	*D* 15.25
Total Column	69	55	56	59	Grand total 239		Grand mean 14.94

The *total* sum of squares is broken up into 4 sums of squares as follows:

Total sum of squares = Sum of squares *between means of columns*
+ Sum of squares *between means of rows*
+ Sum of squares *between means of treatments*
+ Sum of squares for *error*.

This is written as follows:

$$(14.6)\quad \Sigma(x_{ij} - \bar{x})^2 = 4\sum_{c=1}^{4}(\bar{x}_c - \bar{x})^2 + 4\sum_{r=1}^{4}(\bar{x}_r - \bar{x})^2 + 4\sum_{t=1}^{4}(\bar{x}_t - \bar{x})^2 + \sum_{i,j=1}^{4}(x_{ij} - \bar{x}_c - \bar{x}_r - \bar{x}_t + 2\bar{x})^2,$$

where $\bar{x}_c$ represents column means, $\bar{x}_r$ represents row means, and $\bar{x}_t$ represents treatment means. Formula (14.6) can be derived as Formulas (14.1) and (14.5). The estimate of the parent population variance is usually taken to be that of the *error* variance; that is, the variance after the variations due to rows, columns, and treatments have been subtracted from the *total* sum of squares. The following table gives the analysis of variance for the data concerning the growth of oats.

TABLE 14.7

ANALYSIS OF VARIANCE

Source of Variation	Sum of Squares	Degrees of Freedom	Variance
Total....................	130.94	15	8 73
Between means of columns...	30.69	3	10 23
Between means of rows......	11.69	3	3.90
Between means of treatments	72.69	3	24.23
Error....................	15 87	6	2.65

The sum of squares between treatment means is equal to

$$4\Sigma(\bar{x}_t - \bar{x})^2 = \frac{\Sigma T^2}{4} - \frac{(\Sigma x)^2}{16} = \frac{60^2 + 47^2 + 71^2 + 61^2}{4} - \frac{(239)^2}{16} = 72.69.$$

There is a significant difference between treatment means, for the value of F is

$$F = \frac{\text{Treatment variance}}{\text{Error variance}} = \frac{24.23}{2.65} = 9.14,$$

which according to Table III is too large to be due to random sampling. The experimental error is the square root of the variance for error or is $\sqrt{2.65} = 1.63$. The standard deviation of any treatment mean is

$$\sqrt{\frac{\text{Error variance}}{4}} = \frac{1.63}{2} = 0.815.$$

The difference between 2 treatment means, to be significant, is found, as before, to be 2.0 cm. Hence treatment C is significantly better for growth than the other treatments. Treatments A and C are better than B.

The analysis of variance is better for testing the significance of the difference between 2 means than the test on page 220.

PROBLEMS

1. Analyze the data in the following Latin square by the analysis of variance; where the letters represent different fertilizer treatments and the numbers give yields of a certain variety of corn in pounds.

N 72	P 64	K 54	*C* 60	*M* 78
C 56	K 52	*M* 80	P 61	N 71
M 75	N 67	*C* 53	K 58	P 60
P 57	*C* 58	N 70	*M* 83	K 62
K 55	*M* 73	P 54	N 66	*C* 59

where N represents a nitrogen fertilizer, P a phosphorus fertilizer, K a potassium fertilizer, *C* the control and *M* manure.

2. The following data are grades of potatoes from three states judged as to desirability:

State *A*	State *B*	State *C*	State *A*	State *B*	State *C*
4 2	3.8	4.0	4.6	4 0	3.9
3 4	3 5	4.1	2.7	3.2	4.8
4 9	4.0	3.7	3.4	3 6	4.3
3 0	4 1	3.1	3 9	4.0	4 5
3 6	2 9	3.6	3.0	3 1	3.7
4.6	3 9	3.3	3 7	2 8	3 9
4 4	3 7	4.0	3 5	4.0	4 2
3.3	2 8	3.9	2.9	2.8	3.8
3.7	3.0	3.9	3.9	3.0	3 8
3.5	3.4	3.8	3.4	3.6	4.3

By the analysis of variance, test whether or not there are significant differences between the means.

3. The following data pertain to fertilizers and varieties of a certain crop. Test for significant differences between means of varieties and means of yields on different fertilizers.

Yields in Grams

	Varieties			
	A	*B*	*C*	*D*
Fertilizer I.	36	42	34	34
Fertilizer II	38	39	29	39
Fertilizer III.	30	44	28	32

CHAPTER 15

STANDARD ERRORS OF CERTAIN STATISTICS

STANDARD ERROR OF THE MEAN

In Chapter 11 the standard deviation of the distribution made up of all sample averages was found to be

$$(15.1) \qquad \sigma_{\text{mean}} = \cdot \frac{\sqrt{\dfrac{\Sigma(v - M_v)^2}{n - 1}}}{\sqrt{n}} = \frac{s_v}{\sqrt{n}}$$

where s_v is the best estimate of the standard deviation of the parent from which the sample came and n is number of items in the sample.

Example. The mean and value of s_v for the weights of 25 steers in a feeding experiment are 542.3 pounds and 22.1 pounds respectively. According to (15.1), the standard deviation of the distribution made up of all possible averages of the weights of 25 steers similar to the above is

$$\sigma_{\text{mean}} = \frac{22\ 1}{\sqrt{25}} = 4.42 \text{ lb.};$$

this means that the probability of the true mean lying between (542.3 − 4.42 = 537.88) and (542.3 + 4 42 = 546.72) is about 0.682.

The standard deviation of the distribution made up of all sample averages is called the standard error of the mean, and shows how accurate the mean of the sample under consideration is. The following example will illustrate the use of the standard error of the mean.

The following information is known concerning the average weight of the chicks of the first and third brood of a certain hen.

	Mean Weight at 9 Weeks	σ_{mean}
First brood	549 2 grams	5 1 grams
Third brood	421 7 "	4.6 "
Difference.	127 5 "	

Are the two averages significantly different?

The test for significance, as given on page 220, is

$$\text{Significance} = \frac{M_1 - M_2}{\text{Standard deviation of the difference of the means}}$$

$$= \frac{549.2 - 421.7}{\sqrt{(5.1)^2 + (4.6)^2}} = \frac{127.5}{\sqrt{47.17}} > 2.6,$$

which shows that there is a significant difference between the average weight of the first and third brood at 9 weeks of age.

STANDARD ERROR OF A PERCENTAGE

In Chapter 7 it was shown that the standard deviation of a Bernoulli distribution is

$$\sigma = \sqrt{npq}, \tag{15.2}$$

where n is the number of independent trials, p the probability of a success in each trial, and q the probability of a failure. This will be used to derive the standard error of a percentage.

Let N items be divided into classes with frequencies, f_1'', f_2'', f_3'', . . . , f_k'', and let n items be drawn at random from the population of N items. The probability of getting an item from the class with frequency f_i'' is $f_i''/N = p$, and the probability of getting no item from this class is $1 - f_i''/N = q$. The probabilities of getting the various numbers from 0 to n from this class are given by terms of $(q + p)^n$. The standard deviation of this distribution according to (15.2) is

$$\sigma = \sqrt{npq} = \sqrt{n\frac{f_i''}{N}\left(1 - \frac{f_i''}{N}\right)}. \tag{15.3}$$

Let $f_i' = n \cdot \dfrac{f_i''}{N}$; this is the number of items which one can

expect to come from the class with frequency f_i'' in n random drawings. Substituting this in (15 3) gives

$$\text{(15.4)} \qquad \sigma = \sqrt{f_i'\left(1 - \frac{f_i'}{n}\right)} = \sqrt{n\left(\frac{f_i'}{n}\right)\left(1 - \frac{f_i'}{n}\right)}.$$

In practice N and f_i'' are never known and hence f_i' is not known. If a drawing of n is made from the population the number of items, f_i, coming from the ith class will usually lie within $f_i \pm 3\sigma$; hence if σ is small f_i' may be replaced by f_i, the observed frequency for the class under consideration. Formula (15.4) becomes

$$\text{(15.5)} \qquad \sigma_{f_i} = \sqrt{n\left(\frac{f_i}{n}\right)\left(1 - \frac{f_i}{n}\right)},$$

which is considered to be the standard error of the frequency f_i. It is necessary to prove a theorem before the standard error of the percentage f_i/n can be obtained.

THEOREM 15 1. Let the variates of a given distribution be multiplied by a constant, thus forming a new distribution. The mean and the standard deviation of the new distribution are equal respectively to the mean and standard deviation of the given distribution multiplied by the constant.

PROOF: Let the given distribution be given by the first and third columns in the following table and the new distribution be represented by the second and third columns, where k is the constant.

VARIATES	VARIATES	FREQUENCIES
v	kv	f
v_1	kv_1	f_1
v_2	kv_2	f_2
v_3	kv_3	f_3
...	...	...
v_n	kv_n	f_n

The mean of the new distribution is

$$M_{kv} = \frac{kv_1f_1 + kv_2f_2 + \ldots + kv_nf_n}{f_1 + f_2 + \ldots + f_n} = k\frac{\Sigma v \cdot f}{\Sigma f} = kM_v.$$

The standard deviation of the new distribution is

$$\sigma_{kv} = \sqrt{\frac{\Sigma(kv - kM_v)^2}{\Sigma f}} = \sqrt{\frac{k^2\Sigma(v - M_v)^2}{\Sigma f}} = k\sigma_v.$$

The standard deviation of the frequency of a class was found to be (15.5). By employing the results of the above theorem and by allowing $k = 1/n$, the standard deviation of the percentage $f_i/n = p$ is

$$\sigma_p = \frac{1}{n}\sqrt{n\left(\frac{f_i}{n}\right)\left(1 - \frac{f_i}{n}\right)} = \sqrt{\frac{pq}{n}}. \tag{15.6}$$

The following example will illustrate the use of formula (15.6). Community A, of population 1,500, contains 30 blind people, and community B, of population 2,000, contains 45 blind; is there a significant difference between the percentages of blind in the two communities?

According to the above formula the standard error of the first percentage, $p_1 = 30/1{,}500 = 1/50$, is

$$\sigma_{p_1} = \sqrt{\frac{1}{50}\cdot\frac{49}{50}\cdot\frac{1}{1500}} = 0.0036\cdot$$

The standard error of the second percentage $p_2 = 45/2{,}000 = 9/400$ is 0.0016. The test for significance is

$$\text{Significance} = \frac{p_1 - p_2}{\sigma_{(p_1 - p_2)}} = \frac{9/400 - 1/50}{\sqrt{0.0036^2 + 0.0016^2}} = \frac{1}{1.56} < 2.6\cdot$$

Therefore there is no significance between the percentages of blind people in the communities.

THE STANDARD ERROR OF A PRODUCT

When the quantity $A \pm \sigma_A$ is multiplied by the quantity $B \pm \sigma_B$, the question arises as to what is the error of the product. Let the ith measurement of the quantity a be $A_i = A + \bar{x}_i$, where $\bar{x}_i$ is the difference between the measurement and the mean of the measurements; let the jth measurement of the quantity b be $B_j = B + \bar{y}_j$ where $\bar{y}_j$ is the difference between the jth measurement and the mean of all measurements of b. The quantities $\bar{x}_i$ and $\bar{y}_j$ are considered to be errors. Let there be n measurements of a and m measurements of b. The first error of the product

$$A_1B_1 = AB + (B\bar{x}_1 + A\bar{y}_1 + \bar{x}_1\bar{y}_1)$$

is the sum of the quantities in parentheses. The errors of all possible products are listed below.

$$
\begin{array}{llll}
B\bar{x}_1 + A\bar{y}_1 + \bar{x}_1\bar{y}_1, & B\bar{x}_2 + A\bar{y}_1 + \bar{x}_2\bar{y}_1, & \ldots & B\bar{x}_n + A\bar{y}_1 + \bar{x}_n\bar{y}_1 \\
B\bar{x}_1 + A\bar{y}_2 + \bar{x}_1\bar{y}_2, & B\bar{x}_2 + A\bar{y}_2 + \bar{x}_2\bar{y}_2, & \ldots & B\bar{x}_n + A\bar{y}_2 + \bar{x}_n\bar{y}_2 \\
\cdots & \cdots & \cdots & \cdots \\
B\bar{x}_1 + A\bar{y}_m + \bar{x}_1\bar{y}_m, & B\bar{x}_2 + A\bar{y}_m + \bar{x}_2\bar{y}_m, & \ldots & B\bar{x}_n + A\bar{y}_m + \bar{x}_n\bar{y}_m
\end{array}
$$

The square root of the average of the squares of the errors of the product is

$$\sigma_{AB} = \sqrt{\frac{B^2\Sigma\bar{x}^2}{n} + \frac{A^2\Sigma\bar{y}^2}{m} + \frac{2AB\Sigma\bar{x}\cdot\Sigma\bar{y}}{n\cdot m} + \frac{2B\Sigma\bar{x}^2\cdot\Sigma\bar{y}}{n\cdot m} + \frac{2A\Sigma\bar{y}^2\ \Sigma\bar{x}}{n\cdot m} + \frac{\Sigma\bar{x}^2\ \Sigma\bar{y}^2}{n\cdot m}}.$$

The third, fourth, and fifth terms vanish because $\Sigma\bar{x} = \Sigma\bar{y} = 0$, which leaves for the standard error of the product AB

$$\text{(15.7)} \qquad \sigma_{AB} = \sqrt{(B\sigma_A)^2 + (A\sigma_B)^2 + (\sigma_A\sigma_B)^2}.$$

If the quantities σ_A and σ_B are small the product $(\sigma_A\sigma_B)^2$ will be negligible; hence the standard error of the product is

$$\text{(15.8)} \qquad \sigma_{AB} = \sqrt{(B\sigma_A)^2 + (A\sigma_B)^2}.$$

EXAMPLE. The width of a piece of land is 76 ± 0.2 feet; the length is $137 \pm 0\ 4$ feet. According to (15.8), the area of the land is $10{,}412 \pm 41$ square feet.

THE STANDARD ERROR OF A QUOTIENT

Let the quantity a be measured n times with measurements $A_1, A_2, \ldots, A_n$ and the quantity b be measured m times with measurements $B_1, B_2, \ldots, B_m$; and let A and B be the means of the measurements of a and b respectively. Let $\bar{x}_i$ be the deviation of A_i from A, and $\bar{y}_j$ the deviation of B_j from B. The ratio

$$\frac{A_1}{B_1} = \frac{A + \bar{x}_1}{B + \bar{y}_1} = \frac{A + \bar{x}_1}{B + \bar{y}_1} \cdot \frac{B - \bar{y}_1}{B - \bar{y}_1} = \frac{AB - A\bar{y}_1 + B\bar{x}_1 - \bar{x}_1\bar{y}_1}{B^2 - \bar{y}_1^2}$$

$$= \frac{A}{B} + \frac{B\bar{x}_1 - A\bar{y}_1 + (A/B)\bar{y}_1^2 - \bar{x}_1\bar{y}_1}{B^2 - \bar{y}_1^2}.$$

The last fraction is the deviation of the quotient A_1/B_1 of the measurements A_1 and B_1 from the quotient A/B of the means. The standard error of the quotient A/B is the square root of the

average of the squares of all possible errors. The square of the numerator of the first error is

$$B^2\bar{x}_1^2 + A^2\bar{y}_1^2 + \bar{x}_1^2\bar{y}_1^2 + \frac{A^2}{B^2}\bar{y}_1^4 - 2AB\bar{x}_1\bar{y}_1 + 2A\bar{x}_1\bar{y}_1^2$$
$$- 2B\bar{x}_1^2\bar{y}_1 - 2\frac{A^2}{B}\bar{y}_1^3 + 2A\bar{x}_1\bar{y}_1^2 - 2\frac{A}{B}\bar{x}_1\bar{y}_1^3.$$

If the error $\bar{y}_1$ is small, $\bar{y}_1^2$ is much smaller. Let us neglect the quantity $\bar{y}_1^2$ in the denominator. The average of the squares of all possible errors, if $\bar{y}_1^2$ is neglected in the denominators, is

$$\frac{\Sigma e^2}{mn} = \frac{(B\sigma_A)^2 + (A\sigma_B)^2 + \sigma_x^2 \cdot \sigma_y^2 + (A^2/B^2)\dfrac{\Sigma\bar{y}^4}{mn} - 2(A^2/B)\dfrac{\Sigma\bar{y}^3}{mn}}{B^4}.$$

If the last 3 terms in the numerator are negligible, the standard error of the quotient A/B is

$$\sigma_{A/B} = \frac{\sqrt{(B\sigma_A)^2 + (A\sigma_B)^2}}{B^2}, \tag{15.8}$$

where it is understood that B is not zero.

EXAMPLE. The number of calories produced by a given food varies directly as the weight of the food taken. The average of the readings showed the average number of calories to be 63 ± 0.004. The average of the weight readings was $9 \pm 0\ 002$ grams. Find the value of $\dfrac{0.63 \pm 0.004}{0.9 \pm 0.002}$ and its standard error. According to (15.8) the quotient is

$$0.7 \pm \frac{\sqrt{(0.63 \times 0.002)^2 + (0.9 \times 0.004)^2}}{0.9^2} = 0.7 \pm 0.0047 \text{ calorie-gram.}$$

STANDARD ERRORS OF OTHER STATISTICS

Derivations for many of the standard errors of statistics are too complicated to be given in an elementary textbook; the following standard errors are given without derivation. They are based upon a normal parent population and also on a rather large number in the sample.

TABLE 15.1

STANDARD ERRORS OF CERTAIN STATISTICS

Statistics	Standard Error	Statistics	Standard Error
Median	$1.2533\ \sigma/\sqrt{N}$	Rank correlation coefficient	$\frac{1-\rho^2}{\sqrt{N}}[1+0.086\rho^2+0.013\rho^4+0.002\rho^6]$
Standard deviation	$\sigma/\sqrt{2N}$	Multiple correlation coefficient	$\frac{1-(R_{1\,23\,\ldots\,n})^2}{\sqrt{N-n}}$ *
Mean deviation	$\frac{0.6028\ \sigma}{\sqrt{N}}$	Partial correlation coefficient	$\frac{1-(r_{12\,34\,\ldots\,n})^2}{\sqrt{N-2}}$
Coefficient of variation †	$\frac{V}{\sqrt{2N}}\sqrt{1+2\left(\frac{V}{100}\right)^2}$	Semi-interquartile range	$0.7867\ \sigma/\sqrt{N}$
Correlation coefficient	$\frac{1-r^2}{\sqrt{N-2}}$	Regression coefficient $r\cdot\frac{\sigma_y}{\sigma_x}$	$\frac{\sigma_y}{\sigma_x}\sqrt{\frac{1-r_{yx}^2}{N}}$
Either quartile	$1.3626\ \frac{\sigma}{\sqrt{n}}$	Correlation ratio	$\frac{1-\eta^2}{\sqrt{N-m}}$ ‡

* Where n is the number of constants in the regression equations

† $V = \sigma/m$.

‡ m is the number of arrays.

When any statistic is obtained from a set of data its standard error or probable error should always be computed, for this indicates the accuracy of the statistic under consideration. The following problems will enable one to use these standard errors.

PROBLEMS

1. Find the area of the circle and the standard error of the area if the average of 25 measurements of the diameter is 59 ± 0.2 ft.

2. Find the circumference and its standard error of the circle in Problem 1.

3. A cylindrical hole was measured with the following results:

Mean of depth measurements $= 53 \pm 0.3$ ft.,
Mean of diameter measurements $= 7 \pm 0.2$ ft.

Find the volume and its standard error.

4. The following pertain to measurements made on the length and width of a certain rectangular field.

$$\text{Mean of length measurements} = 172 \pm 0.3 \text{ ft.},$$
$$\text{Mean of width measurements} = 81 \pm 0.1 \text{ ft.}$$

Find the area in acres and its standard error in terms of acres.

5. The following data relate to the pressure and volume of a certain gas held at a constant temperature.

$$\text{Mean of readings of the pressure} = 77 \pm 0.1 \text{ cm.},$$
$$\text{Mean of readings of the volume} = 2 \pm 0.2 \text{ cu. cm.}$$

Find the product of the pressure and volume and its standard error.

6. The following pertain to measurements made on a dam:

$$\text{Mean of length measurements} = 374 \pm 0.4 \text{ ft.},$$
$$\text{Mean of width measurements} = 23 \pm 0.2 \text{ ft.},$$
$$\text{Mean of height measurements} = 16 \pm 0.1 \text{ ft.}$$

Find the volume and its standard error.

7. A machine put out 16 imperfect articles in a sample of 500. After the machine was overhauled it put out 3 imperfect articles in a sample of 100. Has the machine been improved?

8. Find the standard errors of the standard deviations found in problem 5 on page 221. Are these standard deviations significantly different?

9. Find the coefficients of variation of the distributions in problem 5 on page 81. Are these significantly different?

10. Find the standard errors of the means of the data pertaining to length of feet given on page 54. Use the t-values in Table III to determine whether or not these means are significantly different.

11. Find the standard error of the correlation coefficient given on page 181.

12. Find the standard error of the rank correlation coefficient given on page 175.

13. The multiple correlation between the number of petals on the second branch terminal flower and the number of petals on the stem terminal flower and the first branch terminal flower is 0.82; this information was obtained from 300 plants on plot A. The multiple correlation coefficient between the number of petals on similar flowers on 200 plants on plot B is 0.74. Are these multiple correlation coefficients significantly different?

14. Partial coefficients of correlation between weight of grain and length of head with length of stalk held constant for barley growing on two study plots are as follows:

$$r_{wh\cdot s} = 0.67 \text{ for 400 plants,}$$
$$r_{wh\cdot s} = 0.60 \text{ for 300 plants,}$$

where w represents weight, h head length, and s length of stalk below the head. Are the partial correlation coefficients significantly different?

15. The correlation ratios between the amount of yellow and red in the color of potatoes from two states are:

$$\text{State } A\text{:}\ \eta_{yr} = 0.89 \text{ from 50 potatoes,}$$
$$\text{State } B\text{:}\ \eta_{yr} = 0.68 \text{ from 73 potatoes.}$$

Are the correlation ratios significantly different?

16. Find the standard error of the regression coefficient which gives the slope of the linear relation between the variables in the example on page 181.

17. Find the standard error of the median of the example on page 41.

18. The average height of one hundred 10-year-old girls is 53 ± 0.3 inches; the average height of one hundred 18-year-old girls is 65 ± 0.4 inches. Are the coefficients of variation significantly different?

19. A zoologist puts each day the same number of white and brown mice in a cage with owls. During the month of September the following was observed.

No. of white mice eaten = 76,
No. of brown mice eaten = 59.

Can one say that the color of the mice determined what the owls ate?

TESTS OF SIGNIFICANCE BETWEEN MEANS OF SMALL SAMPLES

The formulas used for testing significance between means obtained from large samples should not be used when testing significance between means of small samples. "Student" and R. A. Fisher obtained distributions whereby one can make accurate tests between means regardless of the size of the sample. Consider the following example.

Counts were made of the number of pedicels per inflorescence of wild carrots (*Daucus carota*) from two localities A and B.

Locality A No. of Pedicels per Stem Cluster x
50
55
60
47
52
58
46
50
53
54
50
49
44
42
48
42
Total 800

$M_x = 50$ pedicels.
$\Sigma(x - M_x)^2 = 412$.

Locality B No. of Pedicels per Stem Cluster y
47
43
44
48
40
48
39
46
45
43
45
52
Total 540

$M_y = 45$ pedicels.
$\Sigma(y - M_y)^2 = 142$.

Let us combine the sums of the squares of the items from their respective means and divide by the number of degrees of freedom, or the number of items in both sets less 2. The square root of this quantity is called the best estimate of the standard deviation of the parent from which these 2 samples came, or

$$s = \sqrt{\frac{\Sigma(x - M_x)^2 + \Sigma(y - M_y)^2}{n_1 + n_2 - 2}} \tag{15.9}$$

$$= \sqrt{\frac{412 + 142}{26}} = 4.6 \text{ pedicels.}$$

The quantity s is obtained by pooling the sums of the squares of

the items from their respective means. The standard deviations of the means of the sets are respectively

$$\sigma_{M_x} = \frac{s}{\sqrt{16}} = 1.15 \text{ pedicels}, \quad \sigma_{M_y} = \frac{s}{\sqrt{12}} = 1.38.$$

The standard deviation of the difference of the means is

$$\sigma_{\text{Difference of means}} = \sqrt{(\sigma_{M_x})^2 + (\sigma_{M_y})^2}$$

$$= \sqrt{\frac{s^2}{16} + \frac{s^2}{12}} = 1.75^{-}.$$

The value of t is

$$t = \frac{(M_x - M_y) - 0}{\sigma_{\text{Difference of means}}} = \frac{(50 - 45) - 0}{1.75} = 2.86.$$

Look in Table III for t-values at 26 degrees of freedom. The value of t at the 5 per cent point is 2.056, and at the 1 per cent point is 2.779. The value of t obtained above is larger than these two values obtained from the table, hence there is a significant difference between the means. This means that the difference between the means, $(M_x - M_y = 5)$, differs significantly from zero and that the probability that the two samples were taken at random from the same population is very small; from the t-table this probability is less than 0.01.

In general, the value of t is

$$t = \frac{(M_x - M_y) - 0}{s}\sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}}, \tag{15.10}$$

where there are n_1 items in the first set and n_2 items in the second, and s is the value given in (15.9). The number of degrees of freedom at which the t-table is entered is always equal to the denominator in s, that is, $n_1 + n_2 - 2$.

Consider the following data pertaining to percentages of butterfat in milk at 2 different temperatures:

Patron	Low Temperatures 60° – 68° F. x	High Temperatures 85° – 100° F. y	Difference $x - y$
1.	4.03	3.95	0 08
2.	3 54	3.49	0.05
3.	3.35	3 39	−0 04
4	3 81	3.86	−0 05
5.	3 69	3 70	−0 01
6.	3 15	3 20	−0.05
7.	4 11	4.05	0 06
8.	3 51	3 64	−0 13
9.	4.09	4.11	−0 02
10.	3 31	3.40	−0.09
11	3 99	3 94	0.05
12.	3 70	3 74	−0 04
13	3.11	3 14	−0.03
14.	4 90	4 91	−0.01
15.	3 60	3 65	−0.05
16.	3.33	3.44	−0.11
17.	4.33	4.38	−0 05
18	3 90	3.96	−0.06
19.	3.51	3 51	0 00
20.	4.30	4 30	0 00
21.	3 98	3.98	0 00
22.	3.70	3.76	−0 06
23.	3.48	3.48	0 00
24.	3.51	3 58	−0.07
25.	3.94	3.94	0.00
Total	93.87	94.50	−0.63
Mean	3.7548	3.78	−0.0252

Let us apply the above test.

$$s = \sqrt{\frac{4.0776 + 3.738}{48}} = 0.4035.$$

$$t = \frac{0.0252}{0.4035}(3.5355) = 0.221.$$

Entering the t-table at 48 degrees of freedom we see that this test shows no significance between the means, or that the difference between the means does not differ significantly from zero.

If there is no difference between the tests for butterfat at the 2 temperatures the average of the differences of the corresponding readings should be close to zero. The column $x - y$ gives the differences of corresponding readings. The mean of the column of differences is -0.0252. The best estimate of the standard deviation of the parent from which these differences came is

$$s_1 = \sqrt{\frac{\Sigma[(x - y) - (M_{x-y})]^2}{n - 1}},$$

$$s_1 = \sqrt{\frac{0.063024}{24}} = 0.0512.$$

The standard deviation of the mean of the differences is

$$\sigma_{M_{x-y}} = \frac{s_1}{\sqrt{n}} = \frac{0.0512}{5} = 0.01025,$$

where n is the number of differences.
The t-value which "Student" gave is

$$t = \frac{M_{x-y}}{s_{1/\sqrt{n}}} = \frac{-0.0252}{0.01025} = -\ 2.461. \tag{15.11}$$

Entering the t-table at 49 degrees of freedom it is seen that the average of the differences does differ significantly from zero for the 5 per cent point. This contradicts the first test which was used.

The first test cannot be applied when the measurements are correlated. In the case above, the milk for the 2 determinations of butterfat for patron 1 at the 2 temperatures was taken from the same milk can. There is very high correlation between the determinations at low temperatures and high temperatures The first test cannot be applied here because of this high correlation. Another test for determining significance is

$$t = \frac{(M_x - M_y) - 0}{s_2}\sqrt{\frac{n \cdot n}{n + n}}, \tag{15.12}$$

where

$$s_2 = \sqrt{\frac{\Sigma(x-M_x)^2+\Sigma(y-M_y)^2-2\Sigma(x-M_x)(y-M_y)}{n + n - 2}}. \tag{15.13}$$

The quantity s_2 is like s, except the product term,

$$2\Sigma(x - M_x)(y - M_y),$$

has been subtracted from the numerator of s. This takes into consideration the correlation without actually finding the correlation coefficient. Applying this test, the value of s_2 is

$$s_2 = \sqrt{\frac{4.0776 + 3.7380 - 7.7526}{48}} = 0.0360$$

and

$$t = \frac{0.0252 - 0}{0.360}\sqrt{\frac{25 \cdot 25}{50}} = 2.461,$$

which is the same as was obtained in the second test.* The degrees of freedom at which to enter the t-table equal $n - 1$, since this test is the same as that where s_1 is used.

If the correlation coefficient is negative the quantity

$$-2\Sigma(x - M_x)(y - M_y)$$

becomes a positive quantity, and the numerator in the value of s_2 is increased, and hence the value of t is decreased. The last two tests do not apply if the values of x and y are not corresponding values, that is, the x for patron 2 cannot be placed opposite the y value of patron 6.

When there is independence between the sets of variates or measurements the first test is the one to use. The first test can be used to test significance between several sets of independent measurements. The following example concerning length of heads of wheat grown on three different soils, A, B, and C, will illustrate this.

* This last test is the same as the following test: Let

$$S_4 = \sqrt{\sigma_x^2 + \sigma_y^2 - 2r_{xy}\sigma_x\sigma_y};$$

the standard deviation of the difference between the means M_x and M_y is $\frac{S_4}{\sqrt{n}}$ where n is the number of pairs. The t-value is $t = \frac{M_x - M_y}{S_4}\sqrt{n}$.

SOIL A LENGTH OF HEAD	SOIL B LENGTH OF HEAD	SOIL C LENGTH OF HEAD
CM.	CM.	CM.
7 1	6 6	7.2
7.3	6.7	7.2
7.0	6.4	6.9
6.7	6.1	7.4
6.8	6.2	7.1
7 1	6.7	7.6
6 8	6.0	7 1
7 0	5.8	7.5
6.9	5 9	6.7
6 9	6 3	7 3
7.4	6 0	7.2
7 0	6.7	
7 2	6.2	
7 1	6.6	
6.8		
6 9		

Let s_3, similar to the s in (15.9), be

$$s_3 = \sqrt{\frac{\Sigma(x - M_x)^2 + \Sigma(y - M_y)^2 + \Sigma(z - M_z)^2}{n_1 + n_2 + n_3 - 3}}$$

$$= \sqrt{\frac{0.56 + 1.32 + 0.1056}{38}} = 0.229.$$

The standard deviations of the means are

$$\sigma_{M_x} = \frac{s_3}{\sqrt{16}} = 0.057; \quad \sigma_{M_y} = \frac{s_3}{\sqrt{14}} = 0.061$$

$$M_z = \frac{s_3}{\sqrt{11}} = 0.069.$$

The values of t for testing the differences between these means are

$$t = \frac{M_x - M_y}{s_3} \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} = 8.35,$$

$$t = \frac{M_x - M_z}{s_3} \cdot \sqrt{\frac{n_1 \cdot n_3}{n_1 + n_3}} = 2.23,$$

$$t = \frac{M_y - M_z}{s_3} \cdot \sqrt{\frac{n_2 \cdot n_3}{n_2 + n_3}} = 9.75.$$

The above values of t for 38 degrees of freedom show that there are significant differences between the means, or that soil type has much to do with the length of wheat heads.

PROBLEMS

1. The following data pertain to growth of wheat under 2 treatments A and B.

Treatment A	Treatment B
No. of stalks = 34	26
Mean height = 15.6 cm.,	12 1 cm.,
$\Sigma (x - M_x)^2 = 167\ 52.$	$\Sigma (y - M_y)^2 = 145.94.$

Test for significance between the means.

2. The following data pertain to corn planted on 3 different dates.

	April 22	May 1	May 15
No. of stalks.	36	49	64
Mean yield, pounds...	1.7	0 8	0 5
$\Sigma (x - M_x)^2$.........	72.3	78.4	89 6

By the use of s_3 find whether or not there are significant differences between the means.

3. Given the following measurements concerning gain in weights of rats on diets A and B:

Diet A, Gains, Grams	Diet B, Gains, Grams	Diet A, Gains, Grams	Diet B, Gains, Grams
12	21	6	24
16	17	5	27
17	25	16	11
8	30	17	18
29	14	12	23
20	20	15	26
16	16		18

Is it safe to say that one diet is better than the other?

4. Given the following information concerning two sets of yield measurements made on 16 pairs of parallel plots, A and B:

	A	B	
Mean	60 grams.	63 grams.	$r_{AB} = 0.4.$
Standard deviation	3 grams.	4 grams.	

Use the test given in the footnote on page 276 for testing the significance between the means.

5. Suppose there were 9 parallel plots in problem 4 and that the value of $r = -0\ 4$, and the means and standard deviations are as in problem 4. Use the same test for testing significance between means.

6. The following data pertain to heights of first and second sons at 21 years of age and the heights of their fathers at the same age.

Father, Inches	First Son, Inches	Second Son, Inches
66	66.3	66.1
67	66 8	66.9
68	68 5	68 3
69	69 2	69 0
70	70.6	70.4
71	71.8	70 9
72	72 2	71.9
72	71.8	71 5
73	74 0	73 6
74	73.9	73 8
75	76 2	75.7

Test for significance between the means of the heights of the sons by employing s_1 and s_2.

SIGNIFICANCE OF A CORRELATION COEFFICIENT

The value of t for testing whether or not a linear correlation coefficient is significant, that is, whether or not it is significantly different from zero, is

$$t = \frac{r}{\sqrt{1 - r^2}} \cdot \sqrt{n - 2}, \tag{15.14}$$

where r is the linear correlation coefficient and n is the number of pairs of measurements. The number $n - 2$ is the number of degrees of freedom to use for entering the t-table.

Consider the following example.

The correlation coefficient from 27 pairs of measurements is 0.8. The value of t is

$$t = \frac{0.8}{\sqrt{1 - 0.64}} \cdot \sqrt{27 - 2} = 6.67;$$

for 25 degrees of freedom it is seen that $r = 0.8$ is significantly different from zero. This means that the probability that the 27 pairs of measurements came at random from a non-correlated parent is very small, in fact less than 0.01.

PROBLEM

1. Test for significance the correlation coefficient $r = 0.4$, which was determined from 18 pairs of observations.

SIGNIFICANCE BETWEEN STANDARD DEVIATIONS

Table III can be used for testing significance between two standard deviations or two variances for

$$F = \frac{\text{Larger mean square}}{\text{Smaller mean square}} = \frac{\text{Larger Variance}}{\text{Smaller Variance}}$$

$$= \frac{\text{Larger (standard deviation)}^2}{\text{Smaller (standard deviation)}^2}.$$

This will be illustrated by measurements of cherry stems from two branches.

Branch without bud sports	Branch with bud sports
$N = 31,$	$N = 25,$
$\Sigma(x - M_x)^2 = 0.6750,$	$\Sigma(y - M_y)^2 = 1.500,$
$S_x = \sqrt{\frac{0.6750}{30}} = 0.15$ in.	$S_y = \sqrt{\frac{1.500}{24}} = 0.25$ in.

$$F = \frac{0.0625}{0.0225} = 2.78.$$

Entering Table III in the column headed 24 and the row headed 30 we see that the value of F to be significant must be as large or larger than 1.89 at the 5% point or 2.47 at the 1% point.

Hence the two standard deviations are significantly different. Cherries from bud sports have more variability in stem lengths than ordinary cherries.

PROBLEM

1. Use the above method for testing significance between standard deviations found in problem 8 on page 270.

TABLE I

AREA UNDER THE NORMAL CURVE FROM $-\infty$ TO VALUES OF t*

t	A = Area	t	A	t	A	t	A
−4 00	00003						
−3 99	.00003	−3 49	.00024	−2 99	.00140	−2 49	00639
−3 98	.00003	−3 48	.00025+	−2 98	00144	−2 48	00657
−3 97	00004	−3 47	.00026	−2 97	.00149	−2 47	.00676
−3 96	00004	−3 46	.00027	−2 96	.00154	−2 46	.00695−
−3 95	00004	−3 45	.00028	−2 95	00159	−2 45	.00714
−3 94	.00004	−3 44	.00029	−2 94	.00164	−2 44	.00734
−3 93	00004	−3 43	00030	−2 93	00170	−2.43	.00755−
−3 92	00004	−3 42	.00031	−2 92	00175+	−2 42	.00776
−3 91	00005−	−3 41	.00033	−2 91	.00181	−2 41	00798
−3.90	.00005−	−3 40	.00034	−2.90	.00187	−2.40	.00820
−3 89	00005+	−3 39	.00035−	−2 89	.00193	−2 39	.00842
−3 88	00005+	−3 38	.00036	−2 88	.00199	−2 38	00866
−3 87	00005+	−3 37	00038	−2.87	00205+	−2 37	00889
−3 86	00006	−3 36	.00039	−2 86	.00212	−2 36	00914
−3.85	.00006	−3 35	00040	−2 85	.00219	−2 35	.00939
−3 84	.00006	−3 34	00042	−2 84	.00226	−2 34	00964
−3 83	00006	−3 33	.00043	−2 83	.00233	−2 33	.00990
−3 82	.00007	−3 32	00045+	−2 82	.00240	−2 32	01017
−3 81	00007	−3 31	00047	−2 81	.00248	−2 31	01044
−3 80	00007	−3.30	00048	−2 80	00256	−2 30	.01072
−3 79	00008	−3 29	.00050	−2 79	.00264	−2 29	01101
−3 78	.00008	−3 28	00052	−2 78	.00272	−2 28	.01130
−3 77	.00008	−3.27	.00054	−2 77	.00280	−2 27	.01160
−3 76	.00009	−3.26	.00056	−2 76	.00289	−2.26	.01191
−3 75	.00009	−3 25	.00058	−2 75	.00298	−2.25	.01222
−3 74	.00009	−3.24	.00060	−2 74	.00307	−2 24	.01255−
−3 73	.00010	−3.23	.00062	−2 73	.00317	−2 23	.01287
−3.72	.00010	−3 22	.00064	−2 72	.00326	−2.22	01321
−3.71	.00010	−3 21	.00066	−2 71	.00336	−2.21	.01355+
−3.70	.00011	−3.20	.00069	−2 70	.00347	−2.20	.01390
−3.69	.00011	−3.19	00071	−2 69	.00357	−2 19	.01426
−3.68	.00012	−3 18	.00074	−2.68	.00368	−2.18	.01463
−3 67	.00012	−3 17	.00076	−2.67	.00379	−2.17	.01500
−3.66	.00013	−3.16	00079	−2 66	.00391	−2.16	.01539
−3.65	.00013	−3.15	00082	−2 65	00403	−2.15	.01578
−3.64	.00014	−3.14	.00085−	−2 64	.00415−	−2.14	.01618
−3 63	.00014	−3.13	00087	−2 63	.00427	−2.13	.01659
−3.62	.00015−	−3.12	.00090	−2.62	.00440	−2.12	.01700
−3 61	.00015+	−3 11	.00094	−2.61	.00453	−2.11	.01743
−3 60	.00016	−3.10	.00097	−2 60	.00466	−2.10	.01786
−3.59	.00017	−3 09	.00100	−2 59	.00480	−2.09	.01831
−3 58	.00017	−3.08	.00104	−2 58	.00494	−2.08	.01876
−3.57	.00018	−3 07	.00107	−2 57	.00509	−2.07	.01923
−3.56	.00019	−3 06	00111	−2.56	.00523	−2.06	.01970
−3 55	.00019	−3 05	.00114	−2.55	.00539	−2.05	.02018
−3.54	.00020	−3 04	.00118	−2.54	.00554	−2 04	.02068
−3.53	.00021	−3 03	.00122	−2.53	00570	−2.03	.02118
−3 52	00022	−3 02	.00126	−2.52	.00587	−2.02	.02169
−3 51	.00022	−3 01	00131	−2 51	.00604	−2 01	.02222
−3.50	00023	−3 00	00135+	−2.50	.00621	−2.00	.02275+

* Tables I and II were taken from L. R. Salvosa's "Tables," *Annals of Mathematical Statistics*, May, 1930. The fifth figure is rounded off from his table with 6 figures.

TABLE I—(*Continued*)

AREA UNDER THE NORMAL CURVE FROM $-\infty$ TO VALUES OF t

t	A	t	A	t	A	t	A
−1 99	02330	−1.49	.06811	− .99	.16109	− 49	31207
−1 98	02385+	−1 48	.06944	− .98	.16354	− 48	31561
−1 97	.02442	−1.47	.07078	− .97	.16602	− .47	31918
−1 96	02500	−1.46	.07215−	− .96	16853	− .46	32276
−1.95	02559	−1.45	.07353	− .95	.17106	− 45	32636
−1 94	.02619	−1.44	.07493	− 94	.17361	− 44	32997
−1.93	02680	−1.43	.07636	− 93	.17619	− .43	33360
−1 92	02743	−1 42	.07780	− 92	.17879	− .42	.33724
−1 91	.02807	−1.41	.07927	− 91	.18141	− .41	34090
−1 90	.02872	−1.40	.08076	− .90	.18406	− .40	34458
−1 89	02938	−1.39	.08226	− .89	.18673	− 39	.34827
−1 88	03005+	−1 38	.08379	− 88	.18943	− .38	.35197
−1.87	03074	−1.37	08534	− .87	.19215+	− 37	.35569
−1 86	03144	−1 36	.08692	− .86	.19490	− .36	.35942
−1 85	03216	−1 35	.08851	− .85	19766	− 35	.36317
−1.84	.03288	−1.34	09012	− .84	.20045+	− 34	.36693
−1 83	.03363	−1 33	.09176	− .83	20327	− 33	37070
−1 82	03438	−1.32	.09342	− .82	.20611	− 32	.37448
−1 81	03515−	−1 31	.09510	− .81	.20897	− 31	.37828
−1 80	.03593	−1.30	.09680	− .80	.21186	− 30	.38209
−1 79	03673	−1 29	.09853	− .79	.21476	− 29	.38591
−1.78	.03754	−1 28	.10027	− .78	.21770	− .28	.38974
−1.77	03836	−1.27	.10204	− .77	.22065+	− .27	.39358
−1 76	03920	−1.26	.10384	− .76	.22363	− 26	.39743
−1 75	04006	−1 25	.10565+	− .75	.22663	− .25	40129
−1.74	.04093	−1.24	.10749	− .74	.22965+	− .24	.40517
−1.73	.04182	−1 23	.10935−	− 73	.23270	− .23	40905−
−1.72	.04272	−1 22	.11123	− .72	.23576	− .22	.41294
−1 71	.04363	−1.21	.11314	− .71	.23885+	− .21	.41683
−1.70	.04457	−1.20	11507	− .70	.24196	− 20	.42074
−1.69	.04551	−1 19	.11702	− 69	.24510	− .19	.42466
−1 68	.04648	−1 18	11900	− .68	.24825+	− .18	.42858
−1.67	.04746	−1.17	.12100	− 67	.25143	− .17	.43251
−1.66	.04846	−1.16	12302	− .66	25463	− .16	.43644
−1 65	.04947	−1.15	.12507	− .65	.25785−	− 15	.44038
−1 64	.05050	−1.14	.12714	− .64	26109	− 14	.44433
−1 63	.05155+	−1 13	.12924	− .63	.26435−	− .13	.44828
−1.62	.05262	−1.12	.13136	− .62	.26763	− .12	.45224
−1.61	.05370	−1.11	.13350	− .61	.27093	− .11	.45621
−1.60	.05480	−1.10	.13567	− .60	.27425+	− .10	.46017
−1.59	.05592	−1.09	.13786	− .59	.27760	− .09	.46414
−1 58	.05705+	−1.08	.14007	− .58	.28096	− .08	.46812
−1.57	05821	−1.07	.14231	− .57	.28434	− .07	.47210
−1.56	.05938	−1.06	.14457	− .56	.28774	− .06	.47608
−1.55	.06057	−1.05	.14686	− .55	.29116	− .05	.48006
−1.54	.06178	−1 04	.14917	− .54	.29460	− .04	.48405−
−1.53	.06301	−1 03	.15151	− .53	.29806	− .03	.48803
−1.52	.06426	−1 02	.15386	− .52	.30153	− .02	.49202
−1.51	.06522	−1.01	.15625−	− .51	.30503	− .01	.49601
−1.50	.06681	−1 00	.15866	− .50	.30854	− .00	.50000

TABLE I—(*Continued*)

AREA UNDER THE NORMAL CURVE FROM $-\infty$ TO VALUES OF t

t	A	t	A	t	A	t	A
.00	50000	50	.69146	1.00	84135−	1 50	93319
.01	.50399	.51	.69497	1 01	84375+	1 51	.93448
.02	50798	52	.69847	1.02	.84614	1 52	93575−
.03	51197	.53	.70194	1 03	84850	1 53	93699
.04	51595+	54	.70540	1 04	85083	1 54	93822
.05	.51994	.55	.70884	1 05	85314	1 55	93943
06	52392	.56	.71226	1 06	85543	1 56	.94062
.07	52790	57	.71566	1 07	85769	1 57	94179
.08	.53188	58	.71904	1 08	.85993	1 58	94295−
.09	53586	.59	72241	1 09	.86214	1 59	.94408
.10	53983	60	.72575−	1 10	.86433	1 60	94520
.11	54380	61	72907	1 11	.86650	1 61	94630
.12	.54776	.62	73237	1.12	.86864	1 62	94738
.13	55172	63	.73565+	1 13	.87076	1 63	94845−
.14	55567	.64	73891	1.14	87286	1 64	94950
.15	55962	.65	74215+	1 15	87493	1 65	95053
16	56356	66	74537	1 16	87698	1 66	95154
.17	56750	.67	.74857	1 17	87900	1 67	.95254
.18	57142	.68	.75175−	1.18	88100	1.68	95352
.19	.57535−	.69	.75490	1 19	.88298	1 69	.95449
.20	57926	.70	75804	1 20	88493	1 70	.95544
.21	.58317	71	.76115−	1 21	.88686	1 71	.95637
.22	.58706	.72	.76424	1 22	88877	1 72	.95728
.23	.59095+	.73	.76731	1 23	89065+	1.73	.95819
.24	.59484	.74	.77035+	1 24	89251	1.74	95907
.25	.59871	.75	.77337	1 25	.89435+	1 75	.95994
.26	.60257	.76	.77637	1.26	89617	1 76	96080
.27	.60642	.77	.77935+	1.27	89796	1 77	.96164
.28	.61026	.78	.78231	1.28	.89973	1.78	.96246
.29	.61409	.79	.78524	1.29	.90148	1 79	96327
.30	.61791	.80	.78815−	1.30	.90320	1 80	.96407
.31	.62172	.81	.79103	1 31	.90490	1 81	.96485+
.32	62552	.82	.79389	1 32	.90658	1 82	.96562
.33	.62930	.83	.79673	1.33	.90824	1.83	.96638
.34	.63307	.84	.79955−	1 34	.90988	1 84	.96712
.35	.63683	.85	.80234	1 35	.91149	1 85	.96784
.36	.64058	.86	.80511	1 36	.91309	1 86	.96856
.37	.64431	.87	.80785+	1 37	.91466	1 87	.96926
.38	.64803	.88	.81057	1 38	.91621	1.88	.96995−
.39	.65173	.89	.81327	1.39	.91774	1 89	.97062
.40	.65542	90	.81594	1.40	.91924	1.90	.97128
.41	.65910	.91	.81859	1 41	92073	1 91	.97193
.42	.66276	.92	.82121	1.42	.92220	1 92	.97257
.43	.66640	93	.82381	1 43	.92364	1 93	.97320
.44	.67003	.94	.82639	1.44	.92507	1.94	.97381
.45	.67365−	.95	.82894	1.45	.92647	1 95	.97441
.46	.67724	.96	.83147	1.46	.92786	1.96	.97500
.47	.68082	.97	.83398	1.47	.92922	1 97	.97558
.48	.68439	.98	.83646	1.48	.93056	1 98	.97615−
.49	.68793	.99	.83891	1.49	.93189	1.99	.97671

TABLE I—(*Continued*)

AREA UNDER THE NORMAL CURVE FROM $-\infty$ TO VALUES OF t

t	A	t	A	t	A	t	A
2 00	.97725+	2 50	99379	3 00	.99865+	3 50	.99977
2 01	97778	2 51	99396	3 01	.99869	3 51	.99978
2 02	97831	2 52	.99413	3 02	99874	3 52	99978
2 03	.97882	2 53	99430	3 03	.99878	3 53	.99979
2 04	97933	2 54	.99446	3 04	.99882	3 54	.99980
2 05	.97982	2 55	.99461	3 05	.99886	3.55	.99981
2 06	98030	2 56	99477	3 06	.99889	3 56	.99982
2 07	98077	2 57	.99492	3 07	99893	3 57	.99982
2 08	.98124	2.58	.99506	3 08	.99897	3 58	.99983
2.09	98169	2.59	.99520	3 09	.99900	3 59	.99984
2.10	98214	2 60	.99534	3 10	99903	3 60	.99984
2.11	.98257	2 61	.99547	3 11	99907	3.61	.99985−
2.12	.98300	2 62	.99560	3 12	99910	3 62	.99985+
2.13	98341	2 63	99573	3 13	.99913	3 63	.99986
2 14	.98382	2.64	.99586	3 14	.99916	3.64	.99986
2.15	.98422	2.65	.99598	3 15	.99918	3 65	.99987
2.16	.98461	2.66	.99609	3 16	.99921	3 66	.99987
2.17	98500	2 67	.99621	3.17	.99924	3.67	.99988
2.18	.98537	2 68	.99632	3 18	.99926	3.68	.99988
2.19	.98574	2.69	.99643	3.19	.99929	3.69	.99989
2.20	.98610	2 70	.99653	3 20	.99931	3.70	.99989
2.21	98645−	2 71	.99664	3 21	.99934	3.71	.99990
2.22	.98679	2.72	.99674	3 22	.99936	3.72	.99990
2 23	98713	2 73	.99683	3 23	.99938	3.73	.99990
2.24	.98746	2 74	.99693	3 24	.99940	3.74	.99991
2.25	.98778	2 75	.99702	3 25	.99942	3.75	.99991
2 26	.98809	2.76	.99711	3.26	.99944	3.76	.99992
2 27	.98840	2 77	.99720	3.27	.99946	3.77	.99992
2.28	.98870	2 78	.99728	3.28	.99948	3.78	.99992
2.29	.98899	2.79	.99737	3.29	.99950	3.79	.99993
2 30	.98928	2.80	.99745−	3.30	.99952	3.80	.99993
2 31	.98956	2 81	.99752	3 31	.99953	3.81	.99993
2.32	.98983	2 82	.99760	3.32	.99955+	3.82	.99993
2 33	.99010	2 83	.99767	3 33	.99957	3.83	.99994
2.34	.99036	2 84	.99774	3.34	.99958	3.84	.99994
2.35	.99061	2.85	.99781	3.35	.99960	3.85	.99994
2.36	.99086	2 86	.99788	3.36	.99961	3.86	.99994
2.37	.99111	2.87	.99795−	3.37	.99962	3.87	.99995−
2.38	.99134	2.88	.99801	3.38	.99964	3.88	.99995−
2.39	.99158	2.89	.99807	3.39	.99965+	3.89	.99995+
2.40	.99180	2.90	.99813	3.40	.99966	3.90	.99995+
2.41	.99202	2.91	.99819	3.41	.99967	3.91	.99995+
2.42	.99224	2.92	.99825+	3.42	.99969	3.92	.99996
2.43	.99245+	2.93	.99831	3.43	.99970	3.93	.99996
2.44	.99266	2.94	.99836	3.44	.99971	3.94	.99996
2.45	.99286	2.95	.99841	3.45	.99972	3.95	.99996
2.46	.99305+	2.96	.99846	3.46	.99973	3.96	.99996
2.47	.99324	2.97	.99851	3.47	.99974	3.97	.99996
2.48	.99343	2.98	.99856	3.48	.99975−	3.98	.99997
2.49	.99361	2.99	.99861	3.49	.99976	3.99	.99997
						4.00	.99997
						4.50	.999997

TABLE II

Ordinates of the Normal Curve for Values of t

t	Ordinate	t	Ordinate	t	Ordinate	t	Ordinate
−4.00	.00013						
−3 99	00014	−3 49	00090	−2 99	00457	−2 49	01797
−3 98	.00014	−3 48	.00094	−2 98	00471	−2 48	01823
−3 97	00015+	−3 47	00097	−2 97	.00485−	−2 47	01889
−3 96	00016	−3.46	.00100	−2 96	00499	−2 46	01936
−3.95	00016	−3 45	00104	−2 95	00514	−2 45	01984
−3 94	.00017	−3 44	00108	−2.94	00530	−2 44	02033
−3 93	.00018	−3 43	.00111	−2 93	00545+	−2.43	02083
−3 92	00018	−3 42	00115+	−2 92	00562	−2 42	.02134
−3 91	.00019	−3 41	.00119	−2 91	00578+	−2 41	02186
−3.90	.00020	−3.40	.00123	−2 90	.00595+	−2 40	.02240
−3.89	.00021	−3.39	00128	−2 89	.00613	−2.39	02294
−3.88	.00022	−3.38	00132	−2 88	00631	−2.38	02349
−3 87	.00022	−3 37	.00136	−2.87	00649	−2 37	02406
−3 86	.00023	−3 36	.00141	−2 86	.00668	−2 36	02463
−3 85	00024	−3 35	00146	−2.85	.00687	−2 35	02522
−3.84	.00025+	−3 34	00151	−2 84	00707	−2 34	02582
−3 83	.00026	−3 33	00156	−2 83	00727	−2 33	02643
−3 82	.00027	−3 32	.00161	−2 82	00748	−2 32	.02705−
−3 81	00028	−3 31	00167	−2.81	00770	−2 31	.02768
−3 80	00029	−3 30	.00172	−2 80	.00792	−2.30	02833
−3 79	00030	−3 29	.00178	−2.79	00814	−2.29	02899
−3 78	00032	−3 28	00184	−2 78	00837	−2 28	.02966
−3 77	00033	−3 27	00190	−2 77	00861	−2 27	.03034
−3 76	00034	−3 26	00196	−2 76	00885−	−2 26	03103
−3 75	00035+	−3 25	00203	−2 75	00909	−2 25	03174
−3 74	00037	−3 24	00210	−2 74	.00935−	−2.24	.03246
−3 73	00038	−3 23	00217	−2.73	00961	−2 23	.03319
−3 72	.00039	−3.22	00224	−2 72	00987	−2 22	.03394
−3 71	00041	−3 21	.00231	−2.71	01014	−2 21	03470
−3 70	.00043	−3 20	.00238	−2.70	.01042	−2.20	03548
−3 69	.00044	−3.19	00246	−2 69	.01071	−2.19	03626
−3 68	.00046	−3.18	00254	−2.68	01100	−2.18	.03706
−3 67	00047	−3 17	.00262	−2 67	.01130	−2.17	03789
−3 66	.00049	−3 16	.00271	−2.66	01160	−2 16	.03871
−3 65	.00051	−3 15	.00279	−2.65	01191	−2 15	.03955+
−3 64	.00053	−3 14	.00288	−2.64	01223	−2.14	.04041
−3 63	.00055−	−3 13	00298	−2 63	01256	−2 13	.04128
−3.62	.00057	−3.12	.00307	−2 62	.01289	−2.12	.04217
−3 61	.00059	−3.11	.00317	−2.61	01323	−2.11	.04307
−3 60	.00061	−3.10	.00327	−2.60	.01358	−2 10	.04398
−3 59	00063	−3.09	.00337	−2 59	.01394	−2 09	.04492
−3 58	00066	−3 08	.00348	−2 58	01431	−2.08	.04586
−3 57	00068	−3 07	.00358	−2 57	.01468	−2.07	.04682
−3 56	.00071	−3 06	.00370	−2 56	.01506	−2 06	.04780
−3.55	.00073	−3.05	.00381	−2 55	.01545−	−2 05	.04879
−3.54	.00076	−3 04	.00393	−2.54	.01585−	−2 04	.04980
−3 53	.00079	−3 03	.00405−	−2.53	.01625+	−2 03	.05082
−3.52	.00081	−3.02	.00417	−2.52	.01667	−2.02	05186
−3 51	.00084	−3 01	.00430	−2 51	.01710	−2 01	.05292
−3 50	.00087	−3 00	.00443	−2.50	.01753	−2.00	.05399

TABLE II—(*Continued*)

ORDINATES OF THE NORMAL CURVE FOR VALUES OF t

t	Ordinate	t	Ordinate	t	Ordinate	t	Ordinate
−1 99	05508	−1 49	.13147	− 99	.24439	− .49	35381
−1 98	05618	−1 48	.13344	− .98	24681	− 48	35553
−1 97	05730	−1 47	13542	− 97	24923	− .47	35723
−1 96	05844	−1 46	.13742	− 96	25164	− 46	35889
−1 95	05960	−1 45	13943	− 95	25406	− 45	36053
−1 94	06077	−1 44	.14146	− 94	25647	− 44	36124
−1 93	06195+	−1 43	14351	− 93	25888	− 43	36371
−1 92	06316	−1 42	.14556	− 92	26129	− .42	.36526
−1 91	06439	−1 41	14764	− 91	26369	− .41	36678
−1 90	06562	−1 40	14973	− 90	26609	− 40	36827
−1 89	06687	−1 39	15183	− 89	26848	− 39	36973
−1 88	06814	−1 38	15395−	− 88	27086	− 38	37115+
−1 87	06943	−1 37	15608	− .87	27324	− 37	37255−
−1 86	07074	−1 36	15823	− .86	27562	− 36	37391
−1 85	07206	−1 35	16038	− .85	27799	− 35	37524
−1 84	07341	−1 34	.16256	− .84	28034	− .34	.37654
−1 83	07477	−1 33	16474	− .83	28269	− 33	.37780
−1 82	07614	−1 32	16694	− .82	28504	− 32	.37903
−1 81	07754	−1 31	16915−	− .81	.28737	− 31	.38023
−1 80	07895+	−1 30	.17137	− .80	28969	− 30	38139
−1 79	.08038	−1 29	17360	− .79	29200	− .29	.38252
−1 78	08183	−1 28	17585−	− .78	.29431	− 28	.38361
−1 77	08329	−1 27	17810	− 77	29660	− 27	38466
−1 76	.08478	−1 26	18037	− 76	29887	− 26	.38568
−1 75	08628	−1 25	18265−	.75	30114	− .25	.38667
−1 74	08780	−1 24	18494	− 74	30339	− 24	.38762
−1 73	08933	−1 23	.18724	− 73	30563	− .23	38853
−1 72	09089	−1 22	.18954	− 72	30785+	− .22	.38940
−1 71	.09246	−1 21	.19186	− 71	.31006	− .21	39024
−1.70	09405−	−1 20	19419	− 70	.31225+	− .20	.39104
−1.69	.09566	−1 19	19652	− 69	.31443	− .19	.39181
−1.68	09728	−1 18	19886	− 68	.31659	− 18	39253
−1 67	09893	−1 17	20121	− 67	31874	− 17	39322
−1.66	10059	−1 16	20357	− 66	32086	− .16	.39387
−1 65	10227	−1 15	20594	− .65	32297	− .15	39448
−1 64	10396	−1 14	20831	− 64	.32506	− .14	39505+
−1 63	.10568	−1 13	.21069	− .63	32713	− .13	.39559
−1 62	10741	−1 12	21307	− .62	32918	− .12	.39608
−1 61	.10916	−1 11	.21546	− .61	.33122	− .11	.39654
−1.60	.11092	−1 10	.21785+	− 60	.33323	− 10	.39695+
−1.59	.11270	−1 09	22025+	− .59	.33521	− 09	.39733
−1 58	.11451	−1 08	22265+	− 58	33718	− .08	.39767
−1.57	11632	−1 07	22506	− .57	33912	− .07	.39797
−1.56	.11816	−1 06	22747	− 56	.34105−	− .06	.39823
−1.55	.12001	−1.05	22988	− .55	.34294	− .05	.39844
−1.54	.12188	−1.04	23230	− .54	.34482	− 04	.39862
−1.53	.12376	−1 03	.23471	− .53	.34667	− 03	.39876
−1.52	.12567	−1 02	.23713	− .52	.34849	− .02	.39886
−1.51	.12758	−1 01	.23955+	− 51	.35029	− .01	.39892
−1.50	.12952	−1 00	.24197	− .50	.35207	− .00	.39894

TABLE II—(*Continued*)

ORDINATES OF THE NORMAL CURVE FOR VALUES OF t

t	Ordinate	t	Ordinate	t	Ordinate	t	Ordinate
00	39894	50	35207	1 00	.24197	1 50	12952
01	39892	51	35029	1 01	23955+	1 51	12758
.02	39886	52	34849	1 02	23713	1 52	12567
03	39876	53	34667	1 03	23471	1 53	12376
.04	39862	54	34482	1 04	23230	1 54	12188
.05	39844	55	34294	1 05	22988	1 55	12001
.06	39823	56	34105−	1 06	22747	1 56	.11816
.07	39797	57	33912	1 07	22506	1 57	11632
.08	39767	58	33718	1 08	22265+	1 58	11451
.09	39733	59	33521	1 09	22025+	1 59	11270
.10	39695+	60	33323	1 10	21785+	1 60	11092
.11	39654	61	33122	1 11	21546	1 61	10916
.12	39608	62	32918	1 12	21307	1 62	10741
.13	39559	63	32713	1 13	21069	1 63	10568
.14	39505+	64	32506	1 14	20830	1 64	10396
.15	39448	65	32297	1 15	20594	1 65	10227
.16	39387	66	32086	1 16	20357	1 66	10059
.17	39322	67	31874	1 17	20121	1 67	09893
.18	39253	68	31659	1 18	19886	1 68	09728
.19	39181	69	31443	1 19	.19652	1 69	09566
.20	39104	70	31225+	1 20	19419	1 70	.09405−
.21	39024	71	31006	1 21	19186	1 71	09246
.22	38940	72	30785+	1 22	18954	1 72	09089
.23	38853	73	30563	1 23	.18724	1 73	.08933
.24	38762	74	30339	1 24	18494	1 74	08780
25	38667	75	30114	1 25	18265−	1 75	08628
.26	38568	76	29887	1 26	18037	1 76	08478
.27	38466	77	.29660	1 27	17810	1 77	08329
28	38361	78	.29431	1 28	17485−	1 78	08183
.29	38252	79	29200	1 29	17360	1 79	08038
30	38139	80	28969	1 30	17137	1 80	07895+
.31	38023	81	.28737	1 31	16915−	1 81	07754
.32	37903	82	28504	1 32	16694	1 82	07614
.33	37780	83	.28269	1 33	16474	1 83	07477
.34	37654	84	.28034	1 34	16256	1 84	07341
.35	37524	85	.27799	1 35	16038	1 85	07207
.36	.37391	86	.27562	1 36	.15823	1 86	.07074
.37	37255−	87	27324	1 37	15608	1 87	06943
.38	.37115+	88	27086	1 38	15395−	1 88	06814
.39	.36973	89	26848	1 39	15183	1 89	06687
.40	.36827	90	26609	1 40	14973	1 90	06562
.41	.36678	91	26369	1 41	.14764	1 91	06438
.42	.36526	92	26129	1 42	.14556	1 92	06316
.43	.36371	93	.25888	1 43	.14351	1.93	06195+
.44	36214	94	25647	1 44	14146	1 94	06077
.45	.36053	95	25406	1 45	13943	1 95	05960
.46	.35889	96	25164	1.46	13742	1 96	05844
.47	.35723	.97	24923	1 47	13542	1 97	05730
.48	.35553	98	.24681	1 48	.13344	1 98	05618
49	35381	99	24439	1 49	13147	1 99	.05508

TABLE II—(*Continued*)
ORDINATES OF THE NORMAL CURVE FOR VALUES OF *t*

t	Ordinate	*t*	Ordinate	*t*	Ordinate	*t*	Ordinate
2 00	05399	2 50	01753	3 00	00443	3 50	00087
2 01	05292	2 51	01710	3 01	00430	3 51	00084
2 02	05186	2 52	01667	3 02	00417	3 52	00081
2 03	05082	2 53	01625+	3 03	00405−	3 53	00079
2 04	04980	2 54	01585−	3 04	00393	3 54	00076
2 05	04879	2 55	01545−	3 05	00381	3 55	00073
2 06	04780	2 56	01506	3 06	00370	3 56	00071
2 07	04682	2 57	01468	3 07	00358	3 57	00068
2 08	04586	2 58	01431	3 08	00348	3 58	00066
2 09	04492	2 59	01394	3 09	00337	3 59	00063
2 10	04398	2 60	01358	3 10	00327	3 60	00061
2 11	04307	2 61	01323	3 11	00317	3 61	00059
2 12	04217	2 62	01289	3 12	00307	3 62	00057
2.13	04128	2 63	01256	3 13	00298	3 63	00055−
2 14	04041	2 64	01223	3 14	00288	3 64	00053
2 15	03955+	2 65	01191	3 15	00279	3 65	00051
2 16	.03871	2 66	01160	3 16	00271	3 66	00049
2 17	03788	2 67	01130	3 17	.00262	3 67	00047
2 18	03706	2 68	01100	3 18	00254	3 68	00046
2 19	03626	2 69	.01071	3 19	00246	3 69	00044
2.20	03548	2 70	01042	3.20	.00238	3 70	00043
2 21	03470	2 71	01014	3 21	00231	3 71	00041
2 22	03394	2 72	00987	3 22	00224	3 72	00039
2 23	03319	2 73	00961	3 23	00217	3 73	00038
2.24	.03246	2 74	00935−	3 24	.00210	3 74	00037
2 25	03174	2 75	00909	3 25	00203	3 75	00035+
2 26	.03103	2.76	.00885−	3 26	00196	3 76	00034
2 27	03034	2 77	00861	3 27	00190	3 77	.00033
2 28	02966	2 78	00837	3 28	00184	3 78	00032
2 29	.02899	2 79	00814	3 29	00178	3 79	.00030
2 30	02833	2 80	00792	3 30	00172	3 80	00029
2 31	02768	2 81	00770	3 31	00167	3 81	00028
2 32	02705−	2 82	00748	3 32	00161	3 82	00027
2 33	02643	2 83	00727	3 33	00156	3 83	00026
2 34	.02582	2 84	.00707	3 34	.00151	3 84	00025+
2 35	02522	2.85	00687	3 35	00146	3 85	00024
2 36	02463	2 86	00668	3 36	00141	3 86	.00023
2 37	.02406	2 87	00649	3 37	00136	3 87	00022
2 38	02349	2 88	00631	3 38	00132	3 88	00022
2 39	.02294	2 89	.00613	3 39	.00128	3 89	.00021
2.40	.02240	2 90	00595+	3 40	.00123	3 90	.00020
2.41	.02186	2 91	00578	3 41	00119	3 91	00019
2.42	02134	2.92	00562	3 42	.00115+	3 92	00018
2.43	.02083	2 93	.00545+	3 43	.00111	3 93	.00018
2.44	.02033	2 94	00530	3.44	.00108	3 94	.00017
2.45	.01984	2 95	.00514	3 45	.00104	3 95	00016
2.46	.01936	2 96	00499	3.46	00100	3 96	00016
2.47	.01889	2.97	00485−	3 47	00097	3 97	00015+
2.48	.01842	2 98	.00471	3 48	.00094	3 98	00014
2.49	.01797	2.99	00457	3 49	.00090	3 99	.00014
						4 00	.00013
						4.50	.00002

TABLE III*—VALUES OF F AND t

Degrees of Freedom for Smaller Mean Square	Degrees of Freedom for Greater Mean Square: 1	2	3	4	5	6	8	12	24	∞	Values of t
1	161.45 **4052.10**	199 50 **4999 03**	215 72 **5403.49**	224 57 **5625 14**	230 17 **5764.08**	233 97 **5859.39**	238 89 **5981 34**	243 91 **6105 83**	249 04 **6234 16**	254 32 **6366 48**	12 706 **63 657**
2	18.51 **98 49**	19 00 **99 01**	19 16 **99.17**	19 25 **99.25**	19 30 **99.30**	19 33 **99 33**	19 37 **99 36**	19 41 **99 42**	19 45 **99 46**	19 50 **99 50**	4.303 **9 925**
3	10 13 **34 12**	9 55 **30 81**	9 28 **29.46**	9 12 **28 71**	9 01 **28 24**	8 94 **27 91**	8 84 **27 49**	8 74 **27 05**	8 64 **26 60**	8 53 **26 12**	3 182 **5 841**
4	7.71 **21.20**	6 94 **18 00**	6 59 **16.69**	6 39 **15 98**	6 26 **15.52**	6 16 **15 21**	6 04 **14 80**	5 91 **14 37**	5 77 **13 93**	5 63 **13 46**	2 776 **4 604**
5	6 61 **16 26**	5 79 **13 27**	5 41 **12 06**	5 19 **11 39**	5 05 **10 97**	4 95 **10 67**	4 82 **10 27**	4 68 **9 89**	4 53 **9 47**	4 36 **9 02**	2 571 **4 032**
6	5.99 **13.74**	5 14 **10 92**	4 76 **9 78**	4 53 **9 15**	4 39 **8 75**	4 28 **8 47**	4 15 **8 10**	4 00 **7 72**	3 84 **7 31**	3 67 **6 88**	2 447 **3 707**
7	5 59 **12 25**	4 74 **9 55**	4 35 **8 45**	4 12 **7 85**	3 97 **7.46**	3 87 **7 19**	3 73 **6 84**	3 57 **6 47**	3 41 **6 07**	3 23 **5 65**	2 365 **3 499**
8	5 32 **11 26**	4 46 **8 65**	4 07 **7 59**	3 84 **7 01**	3 69 **6 63**	3 58 **6 37**	3 44 **6 03**	3 28 **5 67**	3 12 **5 28**	2 93 **4 86**	2 306 **3 355**
9	5 12 **10.56**	4 26 **8 02**	3.86 **6.99**	3.63 **6 42**	3 48 **6 06**	3 37 **5 80**	3 23 **5 47**	3 07 **5 11**	2 90 **4 73**	2 71 **4 31**	2 262 **3 250**
10	4 96 **10 04**	4 10 **7 56**	3 71 **6 55**	3 48 **5 99**	3 33 **5 64**	3 22 **5 39**	3 07 **5 06**	2 91 **4 71**	2 74 **4 33**	2 54 **3 91**	2 228 **3 169**
11	4 84 **9.65**	3 98 **7.20**	3 59 **6 22**	3 36 **5.67**	3 20 **5.32**	3 09 **5 07**	2 95 **4.74**	2 79 **4.40**	2 61 **4 02**	2 40 **3 60**	2 201 **3 106**

Degrees of Freedom for Smaller Mean Square											
12	4.75 **9.33**	3 88 **6 93**	3 49 **5 95**	3 26 **5 41**	3 11 **5 06**	3 00 **4 82**	2 85 **4 50**	2 69 **4 16**	2.50 **3 78**	2 30 **3.36**	2.179 **3 055**
13	4.67 **9 07**	3 80 **6.70**	3 41 **5.74**	3.18 **5 20**	3 02 **4 86**	2 92 **4.62**	2 77 **4 30**	2 60 **3 96**	2 42 **3 59**	2 21 **3 16**	2 160 **3 012**
14	4 60 **8 86**	3.74 **6.51**	3 34 **5 56**	3 11 **5 03**	2 96 **4 69**	2 85 **4 46**	2 70 **4 14**	2 53 **3 80**	2 35 **3 43**	2 13 **3 00**	2 145 **2 977**
15	4.54 **8 68**	3 68 **6 36**	3 29 **5 42**	3 06 **4 89**	2 90 **4 56**	2.79 **4 32**	2 64 **4 00**	2 48 **3 67**	2 29 **3 29**	2 07 **2 87**	2 131 **2 947**
16	4 49 **8 53**	3 63 **6 23**	3 24 **5 29**	3 01 **4 77**	2 85 **4 44**	2 74 **4 20**	2 59 **3 89**	2 42 **3 55**	2 24 **3 18**	2 01 **2 75**	2 120 **2 921**
17	4.45 **8 40**	3 59 **6.11**	3 20 **5.18**	2 96 **4.67**	2 81 **4 34**	2 70 **4 10**	2 55 **3 79**	2 38 **3 45**	2 19 **3 08**	1 96 **2 65**	2 110 **2 898**
18	4 41 **8 28**	3.55 **6.01**	3 16 **5.09**	2 93 **4 58**	2 77 **4 25**	2 66 **4 01**	2 51 **3 71**	2 34 **3 37**	2 15 **3 10**	1 92 **2 57**	2 101 **2 878**
19	4 38 **8 18**	3 52 **5 93**	3 13 **5 01**	2 90 **4 50**	2.74 **4 17**	2 63 **3 94**	2 48 **3 63**	2 31 **3 30**	2 11 **2 92**	1 88 **2 49**	2 093 **2 861**
20	4.35 **8.10**	3 49 **5.85**	3.10 **4 94**	2 87 **4 43**	2 71 **4 10**	2 60 **3 87**	2 45 **3 56**	2 28 **3 23**	2 08 **2 86**	1 84 **2 42**	2 086 **2 845**
21	4 32 **8 02**	3 47 **5.78**	3 07 **4.87**	2 84 **4.37**	2 68 **4 04**	2 57 **3 81**	2 42 **3 51**	2 25 **3 17**	2 05 **2 80**	1 81 **2 36**	2 080 **2 831**
22	4 30 **7.94**	3 44 **5 72**	3 05 **4 82**	2 82 **4.31**	2 66 **3 99**	2 55 **3 75**	2 40 **3 45**	2 23 **3 12**	2 03 **2 75**	1 78 **2 30**	2 074 **2 819**
23	4 28 **7 88**	3.42 **5 66**	3 03 **4 76**	2 80 **4 26**	2 64 **3 94**	2 53 **3 71**	2 38 **3 41**	2 20 **3 07**	2 00 **2 70**	1 76 **2 26**	2 069 **2 807**

* This table was taken from "Calculation and Interpretation of Analysis of Variance and Covariance," by G W Snedecor, 1934, by permission of the author and the publisher, Collegiate Press, Inc , Ames, Iowa.

TABLE III—(*Continued*)

Degrees of Freedom for Smaller Mean Square	Degrees of Freedom for Greater Mean Square										Values of t
	1	2	3	4	5	6	8	12	24	∞	
24	4.26 **7.82**	3 40 **5 61**	3.01 **4 72**	2 78 **4 22**	2 62 **3 90**	2 51 **3.67**	2 36 **3 36**	2 18 **3.03**	1 98 **2 66**	1 73 **2 21**	2 064 **2 797**
25	4.24 **7.77**	3 38 **5.57**	2 99 **4.68**	2 76 **4.18**	2 60 **3 86**	2 49 **3 63**	2 34 **3.32**	2 16 **2 99**	1 96 **2 62**	1 71 **2 17**	2 060 **2 787**
26	4 22 **7.72**	3 37 **5.53**	2 98 **4 64**	2 74 **4 14**	2 59 **3 82**	2 47 **3.59**	2 32 **3 29**	2 15 **2 96**	1 95 **2 58**	1 69 **2.13**	2 056 **2.779**
27	4.21 **7.68**	3 35 **5.49**	2 96 **4 60**	2 73 **4.11**	2 57 **3.78**	2 46 **3 56**	2 30 **3 26**	2 13 **2 93**	1 93 **2 55**	1 67 **2 10**	2 052 **2 771**
28	4.20 **7 64**	3.34 **5.45**	2.95 **4 57**	2.71 **4 07**	2 56 **3.75**	2 44 **3 53**	2 29 **3 23**	2.12 **2 90**	1 91 **2 52**	1 65 **2.06**	2 048 **2 763**
29	4.18 **7.60**	3.33 **5.42**	2.93 **4.54**	2 70 **4 04**	2 54 **3 73**	2 43 **3 50**	2 28 **3 20**	2 10 **2 87**	1 90 **2.49**	1 64 **2 03**	2 045 **2 756**
30	4 17 **7.56**	3 32 **5.39**	2 92 **4.51**	2 69 **4.02**	2.53 **3 70**	2 42 **3 47**	2 27 **3 17**	2 09 **2.84**	1 89 **2 47**	1 62 **2 01**	2 042 **2.750**
35	4.12 **7.42**	3.27 **5.27**	2.87 **4 40**	2 64 **3 91**	2.48 **3.59**	2.37 **3.37**	2.22 **3.07**	2 04 **2 74**	1 83 **2 37**	1.57 **1.90**	2 030 **2 724**
40	4.08 **7 31**	3.23 **5.18**	2.84 **4 31**	2.61 **3 83**	2.45 **3 51**	2 34 **3 29**	2 18 **2 99**	2.00 **2 66**	1 79 **2 29**	1 52 **1 82**	2 021 **2 704**
45	4.06 **7.23**	3.21 **5.11**	2 81 **4 25**	2 58 **3 77**	2 42 **3 45**	2 31 **3 23**	2 15 **2 94**	1 97 **2 61**	1 76 **2 23**	1 48 **1 75**	2 014 **2 690**
50	4.03 **7.17**	3.18 **5.06**	2.79 **4.20**	2 56 **3.72**	2.40 **3.41**	2 29 **3 19**	2 13 **2 89**	1 95 **2 56**	1 74 **2 18**	1 44 **1 68**	2 008 **2 678**

Degrees of Freedom for Smaller Mean Square											
60	4.00 **7.08**	3.15 **4.98**	2 76 **4 13**	2 52 **3.65**	2 37 **3.34**	2.25 **3.12**	2 10 **2 82**	1 92 **2.50**	1 70 **2 12**	1 39 **1 60**	2 000 **2 660**
70	3.98 **7.01**	3 13 **4 92**	2.74 **4 07**	2 50 **3.60**	2 35 **3 29**	2 23 **3.07**	2 07 **2 78**	1 89 **2.45**	1 67 **2 07**	1 35 **1 43**	1 994 **2 648**
80	3 96 **6.96**	3.11 **4.88**	2 72 **4 04**	2.49 **3.56**	2.33 **3.26**	2.21 **3 04**	2 06 **2.74**	1 88 **2 42**	1 65 **2 03**	1 32 **1.49**	1 990 **2 638**
90	3 95 **6.92**	3 10 **4.85**	2 71 4 01	2.47 3.53	2.32 **3.23**	2 20 **3 01**	2 04 **2.72**	1 86 **2.39**	1 64 **2.00**	1 30 **1 45**	1 987 **2 632**
100	3.94 **6.90**	3 09 **4 82**	2 70 **3 98**	2 46 **3 51**	2 30 **3 21**	2 19 **2 99**	2 03 **2 69**	1 85 **2 37**	1 63 **1.98**	1 28 **1 43**	1 984 **2 626**
125	3 92 **6.84**	3 07 **4.78**	2.68 **3 94**	2 44 **3.47**	2 29 **3 17**	2.17 **2 95**	2 01 **2 66**	1 83 **2 33**	1 60 **1.94**	1 25 **1.37**	1 979 **2 616**
150	3.90 **6.81**	3 06 **4.75**	2 66 **3 91**	2 43 **3 45**	2 27 **3 14**	2 16 **2 92**	2 00 **2.63**	1 82 **2 31**	1 59 **1 92**	1 22 **1 33**	1 976 **2 609**
200	3.89 **6.76**	3 04 **4.71**	2 65 **3.88**	2 42 **3.41**	2.26 **3.11**	2.14 **2 89**	1 98 **2 60**	1.80 **2 28**	1 57 **1 88**	1 19 **1 28**	1 972 **2 601**
300	3 87 **6.72**	3 03 **2 68**	2 64 **3.85**	2 41 **3.38**	2 25 **3 08**	2 13 **2 86**	1 97 **2 57**	1 79 **2 24**	1 55 **1 85**	1 15 **1 22**	1 968 **2 592**
400	3.86 **6.70**	3 02 **4 66**	2 63 **3.83**	2 40 **3.37**	2 24 **3 06**	2 12 **2.85**	1 96 **2.56**	1 78 **2 23**	1 54 **1 84**	1 13 **1 19**	1 966 **2 588**
500	3 86 **6.69**	3 01 **4 65**	2 62 **3 82**	2 39 **3 36**	2 23 **3.05**	2 11 **2.84**	1 96 **2.55**	1 77 **2.22**	1 54 **1 83**	1.11 **1 16**	1 965 **2 586**
1,000	3.85 **6 66**	3 00 **4 63**	2.61 **3 80**	2 38 **3.34**	2 22 **3.04**	2 10 **2 82**	1 95 **2 53**	1 76 **2 20**	1 53 **1.81**	1 08 **1 11**	1 962 **2 581**
∞	3.84 **6.64**	3 00 **4.60**	2 60 **3.78**	2 37 **3 32**	2 21 **3 02**	2 09 **2 80**	1 94 **2 51**	1 75 **2 18**	1 52 **1 78**	1 00 **1 00**	1 960 **2 576**

TABLE IV

Squares, Square Roots, and Reciprocals to 1000*

No.	Square	Square Root	Reciprocal × 100	No.	Square	Square Root	Reciprocal × 100
1	1	1.0000000	100 0000000	51	26 01	7 1414284	1 9607843
2	4	1 4142136	50 0000000	52	27 04	7 2111026	1 9230769
3	9	1 7320508	33.3333333	53	28 09	7 2801099	1 8867925
4	16	2 0000000	25 0000000	54	29 16	7 3484692	1 8518519
5	25	2.2360680	20 0000000	55	30 25	7.4161985	1 8181818
6	36	2 4494897	16 6666667	56	31 36	7.4833148	1 7857143
7	49	2 6457513	14 2857143	57	32 49	7 5498344	1 7543860
8	64	2 8284271	12 5000000	58	33 64	7 6157731	1 7241379
9	81	3 0000000	11 1111111	59	34 81	7 6811457	1 6949153
10	1 00	3.1622777	10.0000000	60	36 00	7.7459667	1 6666667
11	1 21	3 3166248	9 0909091	61	37 21	7 8102497	1 6393443
12	1 44	3 4641016	8 3333333	62	38 44	7 8740079	1 6129032
13	1 69	3 6055513	7 6923077	63	39 69	7 9372539	1 5873016
14	1 96	3 7416574	7.1428571	64	40 96	8 0000000	1 5625000
15	2 25	3 8729833	6.6666667	65	42 25	8 0622577	1 5384615
16	2 56	4 0000000	6.2500000	66	43 56	8.1240384	1 5151515
17	2 89	4 1231056	5 8823529	67	44 89	8.1853528	1 4925373
18	3 24	4.2426407	5.5555556	68	46 24	8 2462113	1 4705882
19	3 61	4 3588989	5 2631579	69	47 61	8 3066239	1 4492754
20	4 00	4.4721360	5.0000000	70	49 00	8.3666003	1 4285714
21	4 41	4.5825757	4 7619048	71	50 41	8 4261498	1 4084507
22	4 84	4 6904158	4 5454545	72	51 84	8 4852814	1 3888889
23	5 29	4 7958315	4 3478261	73	53 29	8 5440037	1 3698630
24	5 76	4.8989795	4 1666667	74	54 76	8.6023253	1 3513514
25	6 25	5 0000000	4.0000000	75	56 25	8.6602540	1.3333333
26	6 76	5 0990195	3 8461538	76	57 76	8 7177979	1.3157895
27	7 29	5 1961524	3 7037037	77	59 29	8 7749644	1 2987013
28	7 84	5 2915026	3.5714286	78	60 84	8.8317609	1 2820513
29	8 41	5 3851648	3 4482759	79	62 41	8 8881944	1 2658228
30	9 00	5.4772256	3 3333333	80	64 00	8.9442719	1.2500000
31	9 61	5 5677644	3 2258065	81	65 61	9.0000000	1 2345679
32	10 24	5.6568542	3.1250000	82	67 24	9.0553851	1 2195122
33	10 89	5.7445626	3 0303030	83	68 89	9.1104336	1 2048193
34	11 56	5.8309519	2 9411765	84	70 56	9.1651514	1.1904762
35	12 25	5.9160798	2.8571429	85	72 25	9.2195445	1.1764706
36	12 96	6 0000000	2 7777778	86	73 96	9 2736185	1 1627907
37	13 69	6.0827625	2.7027027	87	75 69	9.3273791	1 1494253
38	14 44	6 1644140	2 6315789	88	77 44	9.3808315	1 1363636
39	15 21	6 2449980	2.5641026	89	79 21	9.4339811	1 1235955
40	16 00	6.3245553	2.5000000	90	81 00	9.4868330	1 1111111
41	16 81	6.4031242	2.4390244	91	82 81	9.5393920	1 0989011
42	17 64	6.4807407	2.3809524	92	84 64	9.5916630	1 0869565
43	18 49	6.5574385	2.3255814	93	86 49	9.6436508	1.0752688
44	19 36	6 6332496	2 2727273	94	88 36	9.6953597	1.0638298
45	20 25	6 7082039	2.2222222	95	90 25	9 7467943	1.0526316
46	21 16	6 7823300	2 1739130	96	92 16	9.7979590	1.0416667
47	22 09	6 8556546	2 1276596	97	94 09	9 8488578	1 0309278
48	23 04	6 9282032	2.0833333	98	96 04	9 8994949	1 0204082
49	24 01	7 0000000	2.0408163	99	98 01	9 9498744	1.0101010
50	25 00	7 0710678	2.0000000	100	1 00 00	10 0000000	1.0000000

* This table was taken from "Business Statistics," by G. R. Davies and D. Yoder by permission of the authors and John Wiley & Sons, Inc., the publisher.

TABLE IV—*Continued*

SQUARES, SQUARE ROOTS, AND RECIPROCALS TO 1000

No.	Square	Square Root	Reciprocal $\times 10^9$	No.	Square	Square Root	Reciprocal $\times 10^9$
101	1 02 01	10 0498756	9900990	151	2 28 01	12 2882057	6622517
102	1 04 04	10 0995049	9803922	152	2 31 04	12 3288280	6578947
103	1 06 09	10 1488916	9708738	153	2 34 09	12 3693169	6535948
104	1 08 16	10 1980390	9615385	154	2 37 16	12 4096736	6493506
105	1 10 25	10 2469508	9523810	155	2 40 25	12 4498996	6451613
106	1 12 36	10 2956301	9433962	156	2 43 36	12 4899960	6410256
107	1 14 49	10 3440804	9345794	157	2 46 49	12 5299641	6369427
108	1 16 64	10 3923048	9259259	158	2 49 64	12 5698051	6329114
109	1 18 81	10 4403065	9174312	159	2 52 81	12 6095202	6289308
110	1 21 00	10 4880885	9090909	160	2 56 00	12 6491106	6250000
111	1 23 21	10 5356538	9009009	161	2 59 21	12 6885775	6211180
112	1 25 44	10 5830052	8928571	162	2 62 44	12 7279221	6172840
113	1 27 69	10 6301458	8849558	163	2 65 69	12 7671453	6134969
114	1 29 96	10 6770783	8771930	164	2 68 96	12 8062485	6097561
115	1 32 25	10 7238053	8695652	165	2 72 25	12 8452326	6060606
116	1 34 56	10 7703296	8620690	166	2 75 56	12 8840987	6024096
117	1 36 89	10 8166538	8547009	167	2 78 89	12 9228480	5988024
118	1 39 24	10 8627805	8474576	168	2 82 24	12 9614814	5952381
119	1 41 61	10 9087121	8403361	169	2 85 61	13 0000000	5917160
120	1 44 00	10 9544512	8333333	170	2 89 00	13.0384048	5882353
121	1 46 41	11 0000000	8264463	171	2 92 41	13 0766968	5847953
122	1 48 84	11 0453610	8196721	172	2 95 84	13 1148770	5813953
123	1 51 29	11 0905365	8130081	173	2 99 29	13.1529464	5780347
124	1 53 76	11 1355287	8064516	174	3 02 76	13 1909060	5747126
125	1 56 25	11 1803399	8000000	175	3 06 25	13.2287566	5714286
126	1 58 76	11.2249722	7936508	176	3 09 76	13.2664992	5681818
127	1 61 29	11 2694277	7874016	177	3 13 29	13 3041347	5649718
128	1 63 84	11 3137085	7812500	178	3 16 84	13 3416641	5617978
129	1 66 41	11 3578167	7751938	179	3 20 41	13 3790882	5586592
130	1 69 00	11 4017543	7692308	180	3 24 00	13.4164079	5555556
131	1 71 61	11 4455231	7633588	181	3 27 61	13 4536240	5524862
132	1 74 24	11 4891253	7575758	182	3 31 24	13 4907376	5494505
133	1 76 89	11 5325626	7518797	183	3 34 89	13 5277493	5464481
134	1 79 56	11 5758369	7462687	184	3 38 56	13 5646600	5434783
135	1 82 25	11 6189500	7407407	185	3 42 25	13 6014705	5405405
136	1 84 96	11 6619038	7352941	186	3 45 96	13 6381817	5376344
137	1 87 69	11 7046999	7299270	187	3 49 69	13 6747943	5347594
138	1 90 44	11 7473401	7246377	188	3 53 44	13 7113092	5319149
139	1 93 21	11 7898261	7194245	189	3 57 21	13 7477271	5291005
140	1 96 00	11 8321596	7142857	190	3 61 00	13.7840488	5263158
141	1 98 81	11 8743421	7092199	191	3 64 81	13 8202750	5235602
142	2 01 64	11 9163753	7042254	192	3 68 64	13 8564065	5208333
143	2 04 49	11 9582607	6993007	193	3 72 49	13 8924440	5181347
144	2 07 36	12 0000000	6944444	194	3 76 36	13 9283883	5154639
145	2 10 25	12 0415946	6896552	195	3 80 25	13.9642400	5128205
146	2 13 16	12 0830460	6849315	196	3 84 16	14 0000000	5102041
147	2 16 09	12 1243557	6802721	197	3 88 09	14 0356688	5076142
148	2 19 04	12 1655251	6756757	198	3 92 04	14 0712473	5050505
149	2 22 01	12 2065556	6711409	199	3 96 01	14 1067360	5025126
150	2 25 00	12 2474487	6666667	200	4 00 00	14 1421356	5000000

TABLE IV—(*Continued*)

SQUARES, SQUARE ROOTS, AND RECIPROCALS TO 1000

No.	Square	Square Root	Reciprocal $\times 10^9$	No.	Square	Square Root	Reciprocal $\times 10^9$
201	4 04 01	14 1774469	4975124	251	6 30 01	15 8429795	3984064
202	4 08 04	14 2126704	4950495	252	6 35 04	15 8745079	3968254
203	4 12 09	14 2478068	4926108	253	6 40 09	15 9059737	3952569
204	4 16 16	14 2828569	4901961	254	6 45 16	15 9373775	3937008
205	4 20 25	14 3178211	4878049	255	6 50 25	15 9687194	3921569
206	4 24 36	14 3527001	4854369	256	6 55 36	16 0000000	3906250
207	4 28 49	14 3874946	4830918	257	6 60 49	16 0312195	3891051
208	4 32 64	14 4222051	4807692	258	6 65 64	16 0623784	3875969
209	4 36 81	14 4568323	4784689	259	6 70 81	16 0934769	3861004
210	4 41 00	14 4913767	4761905	260	6 76 00	16 1245155	3846154
211	4 45 21	14 5258390	4739336	261	6 81 21	16 1554944	3831418
212	4 49 44	14 5602198	4716981	262	6 86 44	16 1864141	3816794
213	4 53 69	14 5945195	4694836	263	6 91 69	16 2172747	3802281
214	4 57 96	14 6287388	4672897	264	6 96 96	16 2480768	3787879
215	4 62 25	14 6628783	4651163	265	7 02 25	16 2788206	3773585
216	4 66 56	14 6969385	4629630	266	7 07 56	16 3095064	3759398
217	4 70 89	14 7309199	4608295	267	7 12 89	16 3401346	3745318
218	4 75 24	14 7648231	4587156	268	7 18 24	16 3707055	3731343
219	4 79 61	14 7986486	4566210	269	7 23 61	16 4012195	3717472
220	4 84 00	14 8323970	4545455	270	7 29 00	16 4316767	3703704
221	4 88 41	14 8660687	4524887	271	7 34 41	16 4620776	3690037
222	4 92 84	14 8996644	4504505	272	7 39 84	16 4924225	3676471
223	4 97 29	14 9331845	4484305	273	7 45 29	16 5227116	3663004
224	5 01 76	14 9666295	4464286	274	7 50 76	16 5529454	3649635
225	5 06 25	15 0000000	4444444	275	7 56 25	16 5831240	3636364
226	5 10 76	15 0332964	4424779	276	7 61 76	16 6132477	3623188
227	5 15 29	15 0665192	4405286	277	7 67 29	16.6433170	3610108
228	5 19 84	15 0996689	4385965	278	7 72 84	16.6733320	3597122
229	5 24 41	15 1327460	4366812	279	7 78 41	16.7032931	3584229
230	5 29 00	15 1657509	4347826	280	7 84 00	16.7332005	3571429
231	5 33 61	15.1986842	4329004	281	7 89 61	16.7630546	3558719
232	5 38 24	15 2315462	4310345	282	7 95 24	16.7928556	3546099
233	5 42 89	15 2643375	4291845	283	8 00 89	16.8226038	3533569
234	5 47 56	15.2970585	4273504	284	8 06 56	16 8522995	3521127
235	5 52 25	15.3297097	4255319	285	8 12 25	16.8819430	3508772
236	5 56 96	15.3622915	4237288	286	8 17 96	16 9115345	3496503
237	5 61 69	15.3948043	4219409	287	8 23 69	16 9410743	3484321
238	5 66 44	15.4272486	4201681	288	8 29 44	16.9705627	3472222
239	5 71 21	15.4596248	4184100	289	8 35 21	17.0000000	3460208
240	5 76 00	15.4919334	4166667	290	8 41 00	17.0293864	3448276
241	5 80 81	15.5241747	4149378	291	8 46 81	17.0587221	3436426
242	5 85 64	15 5563492	4132231	292	8 52 64	17.0880075	3424658
243	5 90 49	15.5884573	4115226	293	8 58 49	17.1172428	3412969
244	5 95 36	15 6204994	4098361	294	8 64 36	17.1464282	3401361
245	6 00 25	15.6524758	4081633	295	8 70 25	17.1755640	3389831
246	6 05 16	15.6843871	4065041	296	8 76 16	17.2046505	3378378
247	6 10 09	15 7162336	4048583	297	8 82 09	17.2336879	3367003
248	6 15 04	15.7480157	4032258	298	8 88 04	17 2626765	3355705
249	6 20 01	15.7797338	4016064	299	8 94 01	17 2916165	3344482
250	6 25 00	15.8113883	4000000	300	9 00 00	17.3205081	3333333

TABLE IV—(*Continued*)

SQUARES, SQUARE ROOTS, AND RECIPROCALS TO 1000

No.	Square	Square Root	Reciprocal × 10^9	No.	Square	Square Root	Reciprocal × 10^9
301	9 06 01	17 3493516	3322259	351	12 32 01	18 7349940	2849003
302	9 12 04	17 3781472	3311258	352	12 39 04	18 7616630	2840909
303	9 18 09	17 4068952	3300330	353	12 46 09	18 7882942	2832861
304	9 24 16	17 4355958	3289474	354	12 53 16	18 8148877	2824859
305	9 30 25	17 4642492	3278689	355	12 60 25	18 8414437	2816901
306	9 36 36	17 4928557	3267974	356	12 67 36	18 8679623	2808989
307	9 42 49	17 5214155	3257329	357	12 74 49	18 8944436	2801120
308	9 48 64	17 5499288	3246753	358	12 81 64	18 9208879	2793296
309	9 54 81	17 5783958	3236246	359	12 88 81	18 9472953	2785515
310	9 61 00	17 6068169	3225806	360	12 96 00	18 9736660	2777778
311	9 67 21	17 6351921	3215434	361	13 03 21	19 0000000	2770083
312	9 73 44	17 6635217	3205128	362	13 10 44	19 0262976	2762431
313	9 79 69	17 6918060	3194888	363	13 17 69	19 0525589	2754821
314	9 85 96	17 7200451	3184713	364	13 24 96	19 0787840	2747253
315	9 92 25	17 7482393	3174603	365	13 32 25	19 1049732	2739726
316	9 98 56	17 7763888	3164557	366	13 39 56	19 1311265	2732240
317	10 04 89	17 8044938	3154574	367	13 46 89	19.1572441	2724796
318	10 11 24	17 8325545	3144654	368	13 54 24	19 1833261	2717391
319	10 17 61	17 8605711	3134796	369	13 61 61	19 2093727	2710027
320	10 24 00	17.8885438	3125000	370	13 69 00	19.2353841	2702703
321	10 30 41	17 9164729	3115265	371	13 76 41	19 2613603	2695418
322	10 36 84	17 9443584	3105590	372	13 83 84	19 2873015	2688172
323	10 43 29	17 9722008	3095975	373	13 91 29	19 3132079	2680965
324	10 49 76	18 0000000	3086420	374	13 98 76	19 3390796	2673797
325	10 56 25	18 0277564	3076923	375	14 06 25	19 3649167	2666667
326	10 62 76	18 0554701	3067485	376	14 13 76	19 3907194	2659574
327	10 69 29	18 0831413	3058104	377	14 21 29	19 4164878	2652520
328	10 75 84	18.1107703	3048780	378	14 28 84	19 4422221	2645503
329	10 82 41	18 1383571	3039514	379	14 36 41	19 4679223	2638522
330	10 89 00	18.1659021	3030303	380	14 44 00	19 4935887	2631579
331	10 95 61	18 1934054	3021148	381	14 51 61	19 5192213	2624672
332	11 02 24	18 2208672	3012048	382	14 59 24	19 5448203	2617801
333	11 08 89	18 2482876	3003003	383	14 66 89	19 5703858	2610966
334	11 15 56	18 2756669	2994012	384	14 74 56	19 5959179	2604167
335	11 22 25	18 3030052	2985075	385	14 82 25	19 6214169	2597403
336	11 28 96	18 3303028	2976190	386	14 89 96	19 6468827	2590674
337	11 35 69	18 3575598	2967359	387	14 97 69	19 6723156	2583979
338	11 42 44	18 3847763	2958580	388	15 05 44	19 6977156	2577320
339	11 49 21	18 4119526	2949853	389	15 13 21	19 7230829	2570694
340	11 56 00	18 4390889	2941176	390	15 21 00	19 7484177	2564103
341	11 62 81	18 4661853	2932551	391	15 28 81	19 7737199	2557545
342	11 69 64	18 4932420	2923977	392	15 36 64	19 7989899	2551020
343	11 76 49	18 5202592	2915452	393	15 44 49	19.8242276	2544529
344	11 83 36	18 5472370	2906977	394	15 52 36	19.8494332	2538071
345	11 90 25	18.5741756	2898551	395	15 60 25	19.8746069	2531646
346	11 97 16	18 6010752	2890173	396	15 68 16	19.8997487	2525253
347	12 04 09	18.6279360	2881844	397	15 76 09	19 9248588	2518892
348	12 11 04	18 6547581	2873563	398	15 84 04	19 9499373	2512563
349	12 18 01	18 6815417	2865330	399	15 92 01	19 9749844	2506266
350	12 25 00	18 7082869	2857143	400	16 00 00	20 0000000	2500000

TABLE IV—(*Continued*)

Squares, Square Roots, and Reciprocals to 1000

No.	Square	Square Root	Reciprocal $\times 10^9$	No.	Square	Square Root	Reciprocal $\times 10^9$
401	16 08 01	20 0249844	2493766	451	20 34 01	21 2367606	2217295
402	16 16 04	20 0499377	2487562	452	20 43 04	21 2602916	2212389
403	16 24 09	20 0748599	2481390	453	20 52 09	21 2837967	2207506
404	16 32 16	20 0997512	2475248	454	20 61 16	21 3072758	2202643
405	16 40 25	20 1246118	2469136	455	20 70 25	21.3307290	2197802
406	16 48 36	20 1494417	2463054	456	20 79 36	21 3541565	2192982
407	16 56 49	20 1742410	2457002	457	20 88 49	21 3775583	2188184
408	16 64 64	20 1990099	2450980	458	20 97 64	21 4009346	2183406
409	16 72 81	20 2237484	2444988	459	21 06 81	21 4242853	2178649
410	16 81 00	20 2484567	2439024	460	21 16 00	21 4476106	2173913
411	16 89 21	20 2731349	2433090	461	21 25 21	21 4709106	2169197
412	16 97 44	20 2977831	2427184	462	21 34 44	21 4941853	2164502
413	17 05 69	20 3224014	2421308	463	21 43 69	21 5174348	2159827
414	17 13 96	20 3469899	2415459	464	21 52 96	21 5406592	2155172
415	17 22 25	20 3715488	2409639	465	21 62 25	21 5638587	2150538
416	17 30 56	20 3960781	2403846	466	21 71 56	21 5870331	2145923
417	17 38 89	20 4205779	2398082	467	21 80 89	21 6101828	2141328
418	17 47 24	20 4450483	2392344	468	21 90 24	21 6333077	2136752
419	17 55 61	20 4694895	2386635	469	21 99 61	21.6564078	2132196
420	17 64 00	20.4939015	2380952	470	22 09 00	21.6794834	2127660
421	17 72 41	20 5182845	2375297	471	22 18 41	21 7025344	2123142
422	17 80 84	20 5426386	2369668	472	22 27 84	21 7255610	2118644
423	17 89 29	20 5669638	2364066	473	22 37 29	21 7485632	2114165
424	17 97 76	20 5912603	2358491	474	22 46 76	21 7715411	2109705
425	18 06 25	20 6155281	2352941	475	22 56 25	21 7944947	2105263
426	18 14 76	20 6397674	2347418	476	22 65 76	21 8174242	2100840
427	18 23 29	20 6639783	2341920	477	22 75 29	21 8403297	2096436
428	18 31 84	20 6881609	2336449	478	22 84 84	21.8632111	2092050
429	18 40 41	20 7123152	2331002	479	22 94 41	21.8860686	2087683
430	18 49 00	20.7364414	2325581	480	23 04 00	21 9089023	2083333
431	18 57 61	20 7605395	2320186	481	23 13 61	21.9317122	2079002
432	18 66 24	20 7846097	2314815	482	23 23 24	21 9544984	2074689
433	18 74 89	20 8086520	2309469	483	23 32 89	21 9772610	2070393
434	18 83 56	20 8326667	2304147	484	23 42 56	22 0000000	2066116
435	18 92 25	20 8566536	2298851	485	23 52 25	22 0227155	2061856
436	19 00 96	20 8806130	2293578	486	23 61 96	22 0454077	2057613
437	19 09 69	20 9045450	2288330	487	23 71 69	22 0680765	2053388
438	19 18 44	20 9284495	2283105	488	23 81 44	22 0907220	2049180
439	19 27 21	20 9523268	2277904	489	23 91 21	22.1133444	2044990
440	19 36 00	20.9761770	2272727	490	24 01 00	22 1359436	2040816
441	19 44 81	21.0000000	2267574	491	24 10 81	22.1585198	2036660
442	19 53 64	21 0237960	2262443	492	24 20 64	22 1810730	2032520
443	19 62 49	21.0475652	2257336	493	24 30 49	22 2036033	2028398
444	19 71 36	21.0713075	2252252	494	24 40 36	22 2261108	2024291
445	19 80 25	21.0950231	2247191	495	24 50 25	22.2485955	2020202
446	19 89 16	21.1187121	2242152	496	24 60 16	22.2710575	2016129
447	19 98 09	21.1423745	2237136	497	24 70 09	22 2934968	2012072
448	20 07 04	21.1660105	2232143	498	24 80 04	22 3159136	2008032
449	20 16 01	21.1896201	2227171	499	24 90 01	22 3383079	2004008
450	20 25 00	21.2132034	2222222	500	25 00 00	22.3606798	2000000

TABLE IV—(*Continued*)

SQUARES, SQUARE ROOTS, AND RECIPROCALS TO 1000

No.	Square	Square Root	Reciprocal $\times 10^9$	No.	Square	Square Root	Reciprocal $\times 10^9$
501	25 10 01	22 3830293	1996008	551	30 36 01	23 4733892	1814882
502	25 20 04	22 4053565	1992032	552	30 47 04	23.4946802	1811594
503	25 30 09	22 4276615	1988072	553	30 58 09	23 5159520	1808318
504	25 40 16	22 4499443	1984127	554	30 69 16	23 5372046	1805054
505	25 50 25	22 4722051	1980198	555	30 80 25	23 5584380	1801802
506	25 60 36	22 4944438	1976285	556	30 91 36	23 5796522	1798561
507	25 70 49	22 5166605	1972387	557	31 02 49	23 6008474	1795332
508	25 80 64	22 5388553	1968504	558	31 13 64	23 6220236	1792115
509	25 90 81	22 5610283	1964637	559	31 24 81	23 6431808	1788909
510	26 01 00	22.5831796	1960784	560	31 36 00	23.6643191	1785714
511	26 11 21	22 6053091	1956947	561	31 47 21	23 6854386	1782531
512	26 21 44	22 6274170	1953125	562	31 58 44	23 7065392	1779359
513	26 31 69	22 6495033	1949318	563	31 69 69	23 7276210	1776199
514	26 41 96	22 6715681	1945525	564	31 80 96	23 7486842	1773050
515	26 52 25	22.6936114	1941748	565	31 92 25	23 7697286	1769912
516	26 62 56	22 7156334	1937984	566	32 03 56	23 7907545	1766784
517	26 72 89	22 7376340	1934236	567	32 14 89	23 8117618	1763668
518	26 83 24	22 7596134	1930502	568	32 26 24	23.8327506	1760563
519	26 93 61	22.7815715	1926782	569	32 37 61	23 8537209	1757469
520	27 04 00	22 8035085	1923077	570	32 49 00	23 8746728	1754386
521	27 14 41	22 8254244	1919386	571	32 60 41	23 8956063	1751313
522	27 24 84	22 8473193	1915709	572	32 71 84	23 9165215	1748252
523	27 35 29	22.8691933	1912046	573	32 83 29	23.9374184	1745201
524	27 45 76	22.8910463	1908397	574	32 94 76	23 9582971	1742160
525	27 56 25	22 9128785	1904762	575	33 06 25	23 9791576	1739130
526	27 66 76	22 9346899	1901141	576	33 17 76	24 0000000	1736111
527	27 77 29	22 9564806	1897533	577	33 29 29	24 0208243	1733102
528	27 87 84	22.9782506	1893939	578	33 40 84	24 0416306	1730104
529	27 98 41	23 0000000	1890359	579	33 52 41	24 0624188	1727116
530	28 09 00	23.0217289	1886792	580	33 64 00	24 0831892	1724138
531	28 19 61	23 0434372	1883239	581	33 75 61	24 1039416	1721170
532	28 30 24	23 0651252	1879699	582	33 87 24	24 1246762	1718213
533	28 40 89	23 0867928	1876173	583	33 98 89	24.1453929	1715266
534	28 51 56	23.1084400	1872659	584	34 10 56	24 1660919	1712329
535	28 62 25	23.1300670	1869159	585	34 22 25	24.1867732	1709402
536	28 72 96	23 1516738	1865672	586	34 33 96	24 2074369	1706485
537	28 83 69	23 1732605	1862197	587	34 45 69	24.2280829	1703578
538	28 94 44	23 1948270	1858736	588	34 57 44	24 2487113	1700680
539	29 05 21	23 2163735	1855288	589	34 69 21	24 2693222	1697793
540	29 16 00	23.2379001	1851852	590	34 81 00	24 2899156	1694915
541	29 26 81	23 2594067	1848429	591	34 92 81	24 3104916	1692047
542	29 37 64	23 2808935	1845018	592	35 04 64	24 3310501	1689189
543	29 48 49	23 3023604	1841621	593	35 16 49	24 3515913	1686341
544	29 59 36	23.3238076	1838235	594	35 28 36	24 3721152	1683502
545	29 70 25	23.3452351	1834862	595	35 40 25	24 3926218	1680672
546	29 81 16	23.3666429	1831502	596	35 52 16	24.4131112	1677852
547	29 92 09	23 3880311	1828154	597	35 64 09	24 4335834	1675042
548	30 03 04	23 4093998	1824818	598	35 76 04	24.4540385	1672241
549	30 14 01	23.4307490	1821494	599	35 88 01	24.4744765	1669449
550	30 25 00	23.4520788	1818182	600	36 00 00	24 4948974	1666667

TABLE IV—(*Continued*)

Squares, Square Roots, and Reciprocals to 1000

No.	Square	Square Root	Reciprocal × 10^9	No.	Square	Square Root	Reciprocal × 10^9
601	36 12 01	24 5153013	1663894	651	42 38 01	25 5147016	1536098
602	36 24 04	24 5356883	1661130	652	42 51 04	25 5342907	1533742
603	36 36 09	24 5560583	1658375	653	42 64 09	25 5538647	1531394
604	36 48 16	24 5764115	1655629	654	42 77 16	25 5734237	1529052
605	36 60 25	24 5967478	1652893	655	42 90 25	25 5929678	1526718
606	36 72 36	24 6170673	1650165	656	43 03 36	25 6124969	1524390
607	36 84 49	24 6373700	1647446	657	43 16 49	25 6320112	1522070
608	36 96 64	24 6576560	1644737	658	43 29 64	25 6515107	1519757
609	37 08 81	24 6779254	1642036	659	43 42 81	25 6709953	1517451
610	37 21 00	24 6981781	1639344	660	43 56 00	25 6904652	1515152
611	37 33 21	24 7184142	1636661	661	43 69 21	25 7099203	1512859
612	37 45 44	24 7386338	1633987	662	43 82 44	25 7293607	1510574
613	37 57 69	24 7588368	1631321	663	43 95 69	25 7487864	1508296
614	37 69 96	24 7790234	1628664	664	44 08 96	25 7681975	1506024
615	37 82 25	24 7991935	1626016	665	44 22 25	25 7875939	1503759
616	37 94 56	24 8193473	1623377	666	44 35 56	25 8069758	1501502
617	38 06 89	24 8394847	1620746	667	44 48 89	25 8263431	1499250
618	38 19 24	24 8596058	1618123	668	44 62 24	25 8456960	1497006
619	38 31 61	24.8797106	1615509	669	44 75 61	25 8650343	1494768
620	38 44 00	24 8997992	1612903	670	44 89 00	25.8843582	1492537
621	38 56 41	24 9198716	1610306	671	45 02 41	25 9036677	1490313
622	38 68 84	24 9399278	1607717	672	45 15 84	25 9229628	1488095
623	38 81 29	24 9599679	1605136	673	45 29 29	25 9422435	1485884
624	38 93 76	24 9799920	1602564	674	45 42 76	25 9615100	1483680
625	39 06 25	25 0000000	1600000	675	45 56 25	25 9807621	1481481
626	39 18 76	25 0199920	1597444	676	45 69 76	26 0000000	1479290
627	39 31 29	25 0399681	1594896	677	45 83 29	26 0192237	1477105
628	39 43 84	25 0599282	1592357	678	45 96 84	26 0384331	1474926
629	39 56 41	25 0798724	1589825	679	46 10 41	26 0576284	1472754
630	39 69 00	25 0998008	1587302	680	46 24 00	26 0768096	1470588
631	39 81 61	25 1197134	1584786	681	46 37 61	26 0959767	1468429
632	39 94 24	25 1396102	1582278	682	46 51 24	26 1151297	1466276
633	40 06 89	25 1594913	1579779	683	46 64 89	26.1342687	1464129
634	40 19 56	25 1793566	1577287	684	46 78 56	26 1533937	1461988
635	40 32 25	25 1992063	1574803	685	46 92 25	26 1725047	1459854
636	40 44 96	25 2190404	1572327	686	47 05 96	26 1916017	1457726
637	40 57 69	25 2388589	1569859	687	47 19 69	26 2106848	1455604
638	40 70 44	25 2586619	1567398	688	47 33 44	26 2297541	1453488
639	40 83 21	25 2784493	1564945	689	47 47 21	26 2488095	1451379
640	40 96 00	25 2982213	1562500	690	47 61 00	26 2678511	1449275
641	41 08 81	25 3179778	1560062	691	47 74 81	26 2868789	1447178
642	41 21 64	25.3377189	1557632	692	47 88 64	26 3058929	1445087
643	41 34 49	25 3574447	1555210	693	48 02 49	26 3248932	1443001
644	41 47 36	25.3771551	1552795	694	48 16 36	26 3438797	1440922
645	41 60 25	25.3968502	1550388	695	48 30 25	26 3628527	1438849
646	41 73 16	25.4165301	1547988	696	48 44 16	26 3818119	1436782
647	41 86 09	25 4361947	1545595	697	48 58 09	26 4007576	1434720
648	41 99 04	25 4558441	1543210	698	48 72 04	26 4196896	1432665
649	42 12 01	25 4754784	1540832	699	48 86 01	26 4386081	1430615
650	42 25 00	25 4950976	1538462	700	49 00 00	26 4575131	1428571

TABLE IV—(*Continued*)

SQUARES, SQUARE ROOTS, AND RECIPROCALS TO 1000

No.	Square	Square Root	Reciprocal × 10^9	No.	Square	Square Root	Reciprocal × 10^9
701	49 14 01	26 4764046	1426534	751	56 40 01	27 4043792	1331558
702	49 28 04	26 4952826	1424501	752	56 55 04	27.4226184	1329787
703	49 42 09	26 5141472	1422475	753	56 70 09	27 4408455	1328021
704	49 56 16	26 5329983	1420455	754	56 85 16	27 4590604	1326260
705	49 70 25	26 5518361	1418440	755	57 00 25	27.4772633	1324503
706	49 84 36	26 5706605	1416431	756	57 15 36	27 4954542	1322751
707	49 98 49	26 5894716	1414427	757	57 30 49	27 5136330	1321004
708	50 12 64	26 6082694	1412429	758	57 45 64	27 5317998	1319261
709	50 26 81	26 6270539	1410437	759	57 60 81	27 5499546	1317523
710	50 41 00	26 6458252	1408451	760	57 76 00	27 5680975	1315789
711	50 55 21	26 6645833	1406470	761	57 91 21	27 5862284	1314060
712	50 69 44	26 6833281	1404494	762	58 06 44	27 6043475	1312336
713	50 83 69	26 7020598	1402525	763	58 21 69	27 6224546	1310616
714	50 97 96	26 7207784	1400560	764	58 36 96	27 6405499	1308901
715	51 12 25	26.7394839	1398601	765	58 52 25	27 6586334	1307190
716	51 26 56	26 7581763	1396648	766	58 67 56	27 6767050	1305483
717	51 40 89	26 7768557	1394700	767	58 82 89	27 6947648	1303781
718	51 55 24	26 7955220	1392758	768	58 98 24	27 7128129	1302083
719	51 69 61	26 8141754	1390821	769	59 13 61	27 7308492	1300390
720	51 84 00	26 8328157	1388889	770	59 29 00	27 7488739	1298701
721	51 98 41	26 8514432	1386963	771	59 44 41	27 7668868	1297017
722	52 12 84	26 8700577	1385042	772	59 59 84	27 7848880	1295337
723	52 27 29	26 8886593	1383126	773	59 75 29	27 8028775	1293661
724	52 41 76	26 9072481	1381215	774	59 90 76	27 8208555	1291990
725	52 56 25	26 9258240	1379310	775	60 06 25	27.8388218	1290323
726	52 70 76	26 9443872	1377410	776	60 21 76	27 8567766	1288660
727	52 85 29	26 9629375	1375516	777	60 37 29	27 8747197	1287001
728	52 99 84	26 9814751	1373626	778	60 52 84	27 8926514	1285347
729	53 14 41	27 0000000	1371742	779	60 68 41	27 9105715	1283697
730	53 29 00	27 0185122	1369863	780	60 84 00	27.9284801	1282051
731	53 43 61	27 0370117	1367989	781	60 99 61	27 9463772	1280410
732	53 58 24	27 0554985	1366120	782	61 15 24	27 9642629	1278772
733	53 72 89	27 0739727	1364256	783	61 30 89	27 9821372	1277139
734	53 87 56	27.0924344	1362398	784	61 46 56	28 0000000	1275510
735	54 02 25	27.1108834	1360544	785	61 62 25	28 0178515	1273885
736	54 16 96	27 1293199	1358696	786	61 77 96	28 0356915	1272265
737	54 31 69	27 1477439	1356852	787	61 93 69	28 0535203	1270648
738	54 46 44	27.1661554	1355014	788	62 09 44	28 0713377	1269036
739	54 61 21	27.1845544	1353180	789	62 25 21	28 0891438	1267427
740	54 76 00	27.2029410	1351351	790	62 41 00	28.1069386	1265823
741	54 90 81	27.2213152	1349528	791	62 56 81	28 1247222	1264223
742	55 05 64	27.2396769	1347709	792	62 72 64	28 1424946	1262626
743	55 20 49	27.2580263	1345895	793	62 88 49	28.1602557	1261034
744	55 35 36	27.2763634	1344086	794	63 04 36	28 1780056	1259446
745	55 50 25	27.2946881	1342282	795	63 20 25	28.1957444	1257862
746	55 65 16	27 3130006	1340483	796	63 36 16	28 2134720	1256281
747	55 80 09	27.3313007	1338688	797	63 52 09	28 2311884	1254705
748	55 95 04	27.3495887	1336898	798	63 68 04	28.2488938	1253133
749	56 10 01	27 3678644	1335113	799	63 84 01	28 2665881	1251564
750	56 25 00	27 3861279	1333333	800	64 00 00	28.2842712	1250000

TABLE IV—(*Continued*)

SQUARES, SQUARE ROOTS, AND RECIPROCALS TO 1000

No.	Square	Square Root	Reciprocal $\times 10^9$	No.	Square	Square Root	Reciprocal $\times 10^9$
801	64 16 01	28 3019434	1248439	851	72 42 01	29 1719043	1175088
802	64 32 04	28 3196045	1246883	852	72 59 04	29 1890390	1173709
803	64 48 09	28 3372546	1245330	853	72 76 09	29 2061637	1172333
804	64 64 16	28 3548938	1243781	854	72 93 16	29 2232784	1170960
805	64 80 25	28 3725219	1242236	855	73 10 25	29 2403830	1169591
806	64 96 36	28 3901391	1240695	856	73 27 36	29 2574777	1168224
807	65 12 49	28 4077454	1239157	857	73 44 49	29 2745623	1166861
808	65 28 64	28 4253408	1237624	858	73 61 64	29 2916370	1165501
809	65 44 81	28 4429253	1236094	859	73 78 81	29 3087018	1164144
810	65 61 00	28 4604989	1234568	860	73 96 00	29 3257566	1162791
811	65 77 21	28 4780617	1233046	861	74 13 21	29 3428015	1161440
812	65 93 44	28 4956137	1231527	862	74 30 44	29 3598365	1160093
813	66 09 69	28 5131549	1230012	863	74 47 69	29 3768616	1158749
814	66 25 96	28 5306852	1228501	864	74 64 96	29 3938769	1157407
815	66 42 25	28 5482048	1226994	865	74 82 25	29.4108823	1156069
816	66 58 56	28 5657137	1225490	866	74 99 56	29 4278779	1154734
817	66 74 89	28 5832119	1223990	867	75 16 89	29 4448637	1153403
818	66 91 24	28 6006993	1222494	868	75 34 24	29 4618397	1152074
819	67 07 61	28 6181760	1221001	869	75 51 61	29 4788059	1150748
820	67 24 00	28 6356421	1219512	870	75 69 00	29 4957624	1149425
821	67 40 41	28 6530976	1218027	871	75 86 41	29 5127091	1148106
822	67 56 84	28 6705424	1216545	872	76 03 84	29 5296461	1146789
823	67 73 29	28 6879766	1215067	873	76 21 29	29 5465734	1145475
824	67 89 76	28 7054002	1213592	874	76 38 76	29 5634910	1144165
825	68 06 25	28 7228132	1212121	875	76 56 25	29 5803989	1142857
826	68 22 76	28 7402157	1210654	876	76 73 76	29 5972972	1141553
827	68 39 29	28 7576077	1209190	877	76 91 29	29 6141858	1140251
828	68 55 84	28 7749891	1207729	878	77 08 84	29 6310648	1138952
829	68 72 41	28 7923601	1206273	879	77 26 41	29 6479342	1137656
830	68 89 00	28 8097206	1204819	880	77 44 00	29 6647939	1136364
831	69 05 61	28 8270706	1203369	881	77 61 61	29 6816442	1135074
832	69 22 24	28 8444102	1201923	882	77 79 24	29 6984848	1133787
833	69 38 89	28 8617394	1200480	883	77 96 89	29.7153159	1132503
834	69 55 56	28 8790582	1199041	884	78 14 56	29 7321375	1131222
835	69 72 25	28 8963666	1197605	885	78 32 25	29.7489496	1129944
836	69 88 96	28 9136646	1196172	886	78 49 96	29 7657521	1128668
837	70 05 69	28 9309523	1194743	887	78 67 69	29 7825452	1127396
838	70 22 44	28 9482297	1193317	888	78 85 44	29.7993289	1126126
839	70 39 21	28 9654967	1191895	889	79 03 21	29 8161030	1124859
840	70 56 00	28 9827535	1190476	890	79 21 00	29 8328678	1123596
841	70 72 81	29 0000000	1189061	891	79 38 81	29 8496231	1122334
842	70 89 64	29 0172363	1187648	892	79 56 64	29 8663690	1121076
843	71 06 49	29 0344623	1186240	893	79 74 49	29 8831056	1119821
844	71 23 36	29 0516781	1184834	894	79 92 36	29 8998328	1118568
845	71 40 25	29 0688837	1183432	895	80 10 25	29.9165506	1117318
846	71 57 16	29 0860791	1182033	896	80 28 16	29 9332591	1116071
847	71 74 09	29 1032644	1180638	897	80 46 09	29 9499583	1114827
848	71 91 04	29 1204396	1179245	898	80 64 04	29 9666481	1113586
849	72 08 01	29 1376046	1177856	899	80 82 01	29 9833287	1112347
850	72 25 00	29.1547595	1176471	900	81 00 00	30 0000000	1111111

TABLE IV—(*Continued*)

SQUARES, SQUARE ROOTS, AND RECIPROCALS TO 1000

No.	Square	Square Root	Reciprocal × 10^9	No.	Square	Square Root	Reciprocal × 10^9
901	81 18 01	30 0166620	1109878	951	90 44 01	30 8382879	1051525
902	81 36 04	30 0333148	1108647	952	90 63 04	30 8544972	1050420
903	81 54 09	30 0499584	1107420	953	90 82 09	30 8706981	1049318
904	81 72 16	30 0665928	1106195	954	91 01 16	30 8868904	1048218
905	81 90 25	30 0832179	1104972	955	91 20 25	30 9030743	1047120
906	82 08 36	30 0998339	1103753	956	91 39 36	30 9192497	1046025
907	82 26 49	30 1164407	1102536	957	91 58 49	30 9354166	1044932
908	82 44 64	30 1330383	1101322	958	91 77 64	30 9515751	1043841
909	82 62 81	30.1496269	1100110	959	91 96 81	30 9677251	1042753
910	82 81 00	30 1662063	1098901	960	92 16 00	30 9838668	1041667
911	82 99 21	30 1827765	1097695	961	92 35 21	31 0000000	1040583
912	83 17 44	30.1993377	1096491	962	92 54 44	31 0161248	1039501
913	83 35 69	30 2158899	1095290	963	92 73 69	31 0322413	1038422
914	83 53 96	30 2324329	1094092	964	92 92 96	31 0483494	1037344
915	83 72 25	30.2489669	1092896	965	93 12 25	31 0644491	1036269
916	83 90 56	30.2654919	1091703	966	93 31 56	31 0805405	1035197
917	84 08 89	30.2820079	1090513	967	93 50 89	31 0966236	1034126
918	84 27 24	30.2985148	1089325	968	93 70 24	31 1126984	1033058
919	84 45 61	30 3150128	1088139	969	93 89 61	31 1287648	1031992
920	84 64 00	30.3315018	1086957	970	94 09 00	31.1448230	1030928
921	84 82 41	30 3479818	1085776	971	94 28 41	31 1608729	1029866
922	85 00 84	30 3644529	1084599	972	94 47 84	31 1769145	1028807
923	85 19 29	30 3809151	1083424	973	94 67 29	31 1929479	1027749
924	85 37 76	30 3973683	1082251	974	94 86 76	31 2089731	1026694
925	85 56 25	30.4138127	1081081	975	95 06 25	31 2249900	1025641
926	85 74 76	30 4302481	1079914	976	95 25 76	31 2409987	1024590
927	85 93 29	30.4466747	1078749	977	95 45 29	31 2569992	1023541
928	86 11 84	30.4630924	1077586	978	95 64 84	31 2729915	1022495
929	86 30 41	30.4795013	1076426	979	95 84 41	31 2889757	1021450
930	86 49 00	30.4959014	1075269	980	96 04 00	31.3049517	1020408
931	86 67 61	30.5122926	1074114	981	96 23 61	31 3209195	1019368
932	86 86 24	30.5286750	1072961	982	96 43 24	31.3368792	1018330
933	87 04 89	30 5450487	1071811	983	96 62 89	31.3528308	1017294
934	87 23 56	30 5614136	1070664	984	96 82 56	31.3687743	1016260
935	87 42 25	30.5777697	1069519	985	97 02 25	31.3847097	1015228
936	87 60 96	30.5941171	1068376	986	97 21 96	31 4006369	1014199
937	87 79 69	30 6104557	1067236	987	97 41 69	31.4165561	1013171
938	87 98 44	30 6267857	1066098	988	97 61 44	31 4324673	1012146
939	88 17 21	30.6431069	1064963	989	97 81 21	31 4483704	1011122
940	88 36 00	30 6594194	1063830	990	98 01 00	31.4642654	1010101
941	88 54 81	30 6757233	1062699	991	98 20 81	31.4801525	1009082
942	88 73 64	30 6920185	1061571	992	98 40 64	31.4960315	1008065
943	88 92 49	30 7083051	1060445	993	98 60 49	31.5119025	1007049
944	89 11 36	30 7245830	1059322	994	98 80 36	31.5277655	1006036
945	89 30 25	30.7408523	1058201	995	99 00 25	31.5436206	1005025
946	89 49 16	30.7571130	1057082	996	99 20 16	31.5594677	1004016
947	89 68 09	30.7733651	1055966	997	99 40 09	31.5753068	1003009
948	89 87 04	30.7896086	1054852	998	99 60 04	31.5911380	1002004
949	90 06 01	30 8058436	1053741	999	99 80 01	31.6069613	1001001
950	90 25 00	30.8220700	1052632	1000	1 00 00 00	31.6227766	1000000

TABLE V*

Five-place Logarithms: 100–150*

N	0	1	2	3	4	5	6	7	8	9
100	00 000	043	087	130	173	217	260	303	346	389
01	432	475	518	561	604	647	689	732	775	817
02	00 860	903	945	988	*030	*072	*115	*157	*199	*242
03	01 284	326	368	410	452	494	536	578	620	662
04	01 703	745	787	828	870	912	953	995	*036	*078
05	02 119	160	202	243	284	325	366	407	449	490
06	531	572	612	653	694	735	776	816	857	898
07	02 938	979	*019	*060	*100	*141	*181	*222	*262	*302
08	03 342	383	423	463	503	543	583	623	663	703
09	03 743	782	822	862	902	941	981	*021	*060	*100
110	04 139	179	218	258	297	336	376	415	454	493
11	532	571	610	650	689	727	766	805	844	883
12	04 922	961	999	*038	*077	*115	*154	*192	*231	*269
13	05 308	346	385	423	461	500	538	576	614	652
14	05 690	729	767	805	843	881	918	956	994	*032
15	06 070	108	145	183	221	258	296	333	371	408
16	446	483	521	558	595	633	670	707	744	781
17	06 819	856	893	930	967	*004	*041	*078	*115	*151
18	07 188	225	262	298	335	372	408	445	482	518
19	555	591	628	664	700	737	773	809	846	882
120	07 918	954	990	*027	*063	*099	*135	*171	*207	*243
21	08 279	314	350	386	422	458	493	529	565	600
22	636	672	707	743	778	814	849	884	920	955
23	08 991	*026	*061	*096	*132	*167	*202	*237	*272	*307
24	09 342	377	412	447	482	517	552	587	621	656
25	09 691	726	760	795	830	864	899	934	968	*003
26	10 037	072	106	140	175	209	243	278	312	346
27	380	415	449	483	517	551	585	619	653	687
28	10 721	755	789	823	857	890	924	958	992	*025
29	11 059	093	126	160	193	227	261	294	327	361
130	394	428	461	494	528	561	594	628	661	694
31	11 727	760	793	826	860	893	926	959	992	*024
32	12 057	090	123	156	189	222	254	287	320	352
33	385	418	450	483	516	548	581	613	646	678
34	12 710	743	775	808	840	872	905	937	969	*001
35	13 033	066	098	130	162	194	226	258	290	322
36	354	386	418	450	481	513	545	577	609	640
37	672	704	735	767	799	830	862	893	925	956
38	13 988	*019	*051	*082	*114	*145	*176	*208	*239	*270
39	14 301	333	364	395	426	457	489	520	551	582
140	613	644	675	706	737	768	799	829	860	891
41	14 922	953	983	*014	*045	*076	*106	*137	*168	*198
42	15 229	259	290	320	351	381	412	442	473	503
43	534	564	594	625	655	685	715	746	776	806
44	15 836	866	897	927	957	987	*017	*047	*077	*107
45	16 137	167	197	227	256	286	316	346	376	406
46	435	465	495	524	554	584	613	643	673	702
47	16 732	761	791	820	850	879	909	938	967	997
48	17 026	056	085	114	143	173	202	231	260	289
49	319	348	377	406	435	464	493	522	551	580
150	17 609	638	667	696	725	754	782	811	840	869
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

	44	**43**	**42**
1	4.4	4.3	4.2
2	8.8	8.6	8.4
3	13.2	12.9	12.6
4	17.6	17.2	16.8
5	22.0	21.5	21.0
6	26.4	25.8	25.2
7	30.8	30.1	29.4
8	35.2	34.4	33.6
9	39.6	38.7	37.8

	41	**40**	**39**
1	4.1	4	3.9
2	8.2	8	7.8
3	12.3	12	11.7
4	16.4	16	15.6
5	20.5	20	19.5
6	24.6	24	23.4
7	28.7	28	27.3
8	32.8	32	31.2
9	36.9	36	35.1

	38	**37**	**36**
1	3.8	3.7	3.6
2	7.6	7.4	7.2
3	11.4	11.1	10.8
4	15.2	14.8	14.4
5	19.0	18.5	18.0
6	22.8	22.2	21.6
7	26.6	25.9	25.2
8	30.4	29.6	28.8
9	34.2	33.3	32.4

	35	**34**	**33**
1	3.5	3.4	3.3
2	7.0	6.8	6.6
3	10.5	10.2	9.9
4	14.0	13.6	13.2
5	17.5	17.0	16.5
6	21.0	20.4	19.8
7	24.5	23.8	23.1
8	28.0	27.2	26.4
9	31.5	30.6	29.7

	32	**31**	**30**
1	3.2	3.1	3
2	6.4	6.2	6
3	9.6	9.3	9
4	12.8	12.4	12
5	16.0	15.5	15
6	19.2	18.6	18
7	22.4	21.7	21
8	25.6	24.8	24
9	28.8	27.9	27

Prop. Parts

* This table was taken from "Plane Trigonometry with Five-place Tables" by H. A. Simmons and G. D. Gore by permission of the authors and the publisher, John Wiley

TABLE V—(*Continued*)

FIVE-PLACE LOGARITHMS: 150–200

N	0	1	2	3	4	5	6	7	8	9
150	17 609	638	667	696	725	754	782	811	840	869
51	17 898	926	955	984	*013	*041	*070	*099	*127	*156
52	18 184	213	241	270	298	327	355	384	412	441
53	469	498	526	554	583	611	639	667	696	724
54	18 752	780	808	837	865	893	921	949	977	*005
55	19 033	061	089	117	145	173	201	229	257	285
56	312	340	368	396	424	451	479	507	535	562
57	590	618	645	673	700	728	756	783	811	838
58	19 866	893	921	948	976	*003	*030	*058	*085	*112
59	20 140	167	194	222	249	276	303	330	358	385
160	412	439	466	493	520	548	575	602	629	656
61	683	710	737	763	790	817	844	871	898	925
62	20 952	978	*005	*032	*059	*085	*112	*139	*165	*192
63	21 219	245	272	299	325	352	378	405	431	458
64	484	511	537	564	590	617	643	669	696	722
65	21 748	775	801	827	854	880	906	932	958	985
66	22 011	037	063	089	115	141	167	194	220	246
67	272	298	324	350	376	401	427	453	479	505
68	531	557	583	608	634	660	686	712	737	763
69	22 789	814	840	866	891	917	943	968	994	*019
170	23 045	070	096	121	147	172	198	223	249	274
71	300	325	350	376	401	426	452	477	502	528
72	553	578	603	629	654	679	704	729	754	779
73	23 805	830	855	880	905	930	955	980	*005	*030
74	24 055	080	105	130	155	180	204	229	254	279
75	304	329	353	378	403	428	452	477	502	527
76	551	576	601	625	650	674	699	724	748	773
77	24 797	822	846	871	895	920	944	969	993	*018
78	25 042	066	091	115	139	164	188	212	237	261
79	285	310	334	358	382	406	431	455	479	503
180	527	551	575	600	624	648	672	696	720	744
81	25 768	792	816	840	864	888	912	935	959	983
82	26 007	031	055	079	102	126	150	174	198	221
83	245	269	293	316	340	364	387	411	435	458
84	482	505	529	553	576	600	623	647	670	694
85	717	741	764	788	811	834	858	881	905	928
86	26 951	975	998	*021	*045	*068	*091	*114	*138	*161
87	27 184	207	231	254	277	300	323	346	370	393
88	416	439	462	485	508	531	554	577	600	623
89	646	669	692	715	738	761	784	807	830	852
190	27 875	898	921	944	967	989	*012	*035	*058	*081
91	28 103	126	149	171	194	217	240	262	285	307
92	330	353	375	398	421	443	466	488	511	533
93	556	578	601	623	646	668	691	713	735	758
94	28 780	803	825	847	870	892	914	937	959	981
95	29 003	026	048	070	092	115	137	159	181	203
96	226	248	270	292	314	336	358	380	403	425
97	447	469	491	513	535	557	579	601	623	645
98	667	688	710	732	754	776	798	820	842	863
99	29 885	907	929	951	973	994	*016	*038	*060	*081
200	30 103	125	146	168	190	211	233	255	276	298
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

	29	28
1	2.9	2.8
2	5.8	5.6
3	8.7	8.4
4	11.6	11.2
5	14.5	14.0
6	17.4	16.8
7	20.3	19.6
8	23.2	22.4
9	26.1	25.2

	27	26
1	2.7	2.6
2	5.4	5.2
3	8.1	7.8
4	10.8	10.4
5	13.5	13.0
6	16.2	15.6
7	18.9	18.2
8	21.6	20.8
9	24.3	23.4

	25
1	2.5
2	5.0
3	7.5
4	10.0
5	12.5
6	15.0
7	17.5
8	20.0
9	22.5

	24	23
1	2.4	2.3
2	4.8	4.6
3	7.2	6.9
4	9.6	9.2
5	12.0	11.5
6	14.4	13.8
7	16.8	16.1
8	19.2	18.4
9	21.6	20.7

	22	21
1	2.2	2.1
2	4.4	4.2
3	6.6	6.3
4	8.8	8.4
5	11.0	10.5
6	13.2	12.6
7	15.4	14.7
8	17.6	16.8
9	19.8	18.9

Prop. Parts

TABLE V—(*Continued*)
FIVE-PLACE LOGARITHMS: 200–250

N	0	1	2	3	4	5	6	7	8	9
200	30 103	125	146	168	190	211	233	255	276	298
01	320	341	363	384	406	428	449	471	492	514
02	535	557	578	600	621	643	664	685	707	728
03	750	771	792	814	835	856	878	899	920	942
04	30 963	984	*006	*027	*048	*069	*091	*112	*133	*154
05	31 175	197	218	239	260	281	302	323	345	366
06	387	408	429	450	471	492	513	534	555	576
07	597	618	639	660	681	702	723	744	765	785
08	31 806	827	848	869	890	911	931	952	973	994
09	32 015	035	056	077	098	118	139	160	181	201
210	222	243	263	284	305	325	346	366	387	408
11	428	449	469	490	510	531	552	572	593	613
12	634	654	675	695	715	736	756	777	797	818
13	32 838	858	879	899	919	940	960	980	*001	*021
14	33 041	062	082	102	122	143	163	183	203	224
15	244	264	284	304	325	345	365	385	405	425
16	445	465	486	506	526	546	566	586	606	626
17	646	666	686	706	726	746	766	786	806	826
18	33 846	866	885	905	925	945	965	985	*005	*025
19	34 044	064	084	104	124	143	163	183	203	223
220	242	262	282	301	321	341	361	380	400	420
21	439	459	479	498	518	537	557	577	596	616
22	635	655	674	694	713	733	753	772	792	811
23	34 830	850	869	889	908	928	947	967	986	*005
24	35 025	044	064	083	102	122	141	160	180	199
25	218	238	257	276	295	315	334	353	372	392
26	411	430	449	468	488	507	526	545	564	583
27	603	622	641	660	679	698	717	736	755	774
28	793	813	832	851	870	889	908	927	946	965
29	35 984	*003	*021	*040	*059	*078	*097	*116	*135	*154
230	36 173	192	211	229	248	267	286	305	324	342
31	361	380	399	418	436	455	474	493	511	530
32	549	568	586	605	624	642	661	680	698	717
33	736	754	773	791	810	829	847	866	884	903
34	36 922	940	959	977	996	*014	*033	*051	*070	*088
35	37 107	125	144	162	181	199	218	236	254	273
36	291	310	328	346	365	383	401	420	438	457
37	475	493	511	530	548	566	585	603	621	639
38	658	676	694	712	731	749	767	785	803	822
39	37 840	858	876	894	912	931	949	967	985	*003
240	38 021	039	057	075	093	112	130	148	166	184
41	202	220	238	256	274	292	310	328	346	364
42	382	399	417	435	453	471	489	507	525	543
43	561	578	596	614	632	650	668	686	703	721
44	739	757	775	792	810	828	846	863	881	899
45	38 917	934	952	970	987	*005	*023	*041	*058	*076
46	39 094	111	129	146	164	182	199	217	235	252
47	270	287	305	322	340	358	375	393	410	428
48	445	463	480	498	515	533	550	568	585	602
49	620	637	655	672	690	707	724	742	759	777
250	39 794	811	829	846	863	881	898	915	933	950
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

	22	**21**
1	2.2	2.1
2	4.4	4.2
3	6.6	6.3
4	8.8	8.4
5	11.0	10.5
6	13.2	12.6
7	15.4	14.7
8	17.6	16.8
9	19.8	18.9

	20
1	2
2	4
3	6
4	8
5	10
6	12
7	14
8	16
9	18

	19
1	1.9
2	3.8
3	5.7
4	7.6
5	9.5
6	11.4
7	13.3
8	15.2
9	17.1

	18
1	1.8
2	3.6
3	5.4
4	7.2
5	9.0
6	10.8
7	12.6
8	14.4
9	16.2

	17
1	1.7
2	3.4
3	5.1
4	6.8
5	8.5
6	10.2
7	11.9
8	13.6
9	15.3

Prop. Parts

TABLE V—(*Continued*)
FIVE-PLACE LOGARITHMS: 250–300

N	0	1	2	3	4	5	6	7	8	9
250	39 794	811	829	846	863	881	898	915	933	950
51	39 967	985	*002	*019	*037	*054	*071	*088	*106	*123
52	40 140	157	175	192	209	226	243	261	278	295
53	312	329	346	364	381	398	415	432	449	466
54	483	500	518	535	552	569	586	603	620	637
55	654	671	688	705	722	739	756	773	790	807
56	824	841	858	875	892	909	926	943	960	976
57	40 993	*010	*027	*044	*061	*078	*095	*111	*128	*145
58	41 162	179	196	212	229	246	263	280	296	313
59	330	347	363	380	397	414	430	447	464	481
260	497	514	531	547	564	581	597	614	631	647
61	664	681	697	714	731	747	764	780	797	814
62	830	847	863	880	896	913	929	946	963	979
63	41 996	*012	*029	*045	*062	*078	*095	*111	*127	*144
64	42 160	177	193	210	226	243	259	275	292	308
65	325	341	357	374	390	406	423	439	455	472
66	488	504	521	537	553	570	586	602	619	635
67	651	667	684	700	716	732	749	765	781	797
68	813	830	846	862	878	894	911	927	943	959
69	42 975	991	*008	*024	*040	*056	*072	*088	*104	*120
270	43 136	152	169	185	201	217	233	249	265	281
71	297	313	329	345	361	377	393	409	425	441
72	457	473	489	505	521	537	553	569	584	600
73	616	632	648	664	680	696	712	727	743	759
74	775	791	807	823	838	854	870	886	902	917
75	43 933	949	965	981	996	*012	*028	*044	*059	*075
76	44 091	107	122	138	154	170	185	201	217	232
77	248	264	279	295	311	326	342	358	373	389
78	404	420	436	451	467	483	498	514	529	545
79	560	576	592	607	623	638	654	669	685	700
280	716	731	747	762	778	793	809	824	840	855
81	44 871	886	902	917	932	948	963	979	994	*010
82	45 025	040	056	071	086	102	117	133	148	163
83	179	194	209	225	240	255	271	286	301	317
84	332	347	362	378	393	408	423	439	454	469
85	484	500	515	530	545	561	576	591	606	621
86	637	652	667	682	697	712	728	743	758	773
87	788	803	818	834	849	864	879	894	909	924
88	45 939	954	969	984	*000	*015	*030	*045	*060	*075
89	46 090	105	120	135	150	165	180	195	210	225
290	240	255	270	285	300	315	330	345	359	374
91	389	404	419	434	449	464	479	494	509	523
92	538	553	568	583	598	613	627	642	657	672
93	687	702	716	731	746	761	776	790	805	820
94	835	850	864	879	894	909	923	938	953	967
95	46 982	997	*012	*026	*041	*056	*070	*085	*100	*114
96	47 129	144	159	173	188	202	217	232	246	261
97	276	290	305	319	334	349	363	378	392	407
98	422	436	451	465	480	494	509	524	538	553
99	567	582	596	611	625	640	654	669	683	698
300	47 712	727	741	756	770	784	799	813	828	842
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

	18	**17**	**16**	**15**	**14**
1	1.8	1.7	1.6	1.5	1.4
2	3.6	3.4	3.2	3.0	2.8
3	5.4	5.1	4.8	4.5	4.2
4	7.2	6.8	6.4	6.0	5.6
5	9.0	8.5	8.0	7.5	7.0
6	10.8	10.2	9.6	9.0	8.4
7	12.6	11.9	11.2	10.5	9.8
8	14.4	13.6	12.8	12.0	11.2
9	16.2	15.3	14.4	13.5	12.6

TABLE V—(*Continued*)

FIVE-PLACE LOGARITHMS: 300–350

N	0	1	2	3	4	5	6	7	8	9
300	47 712	727	741	756	770	784	799	813	828	842
01	47 857	871	885	900	914	929	943	958	972	986
02	48 001	015	029	044	058	073	087	101	116	130
03	144	159	173	187	202	216	230	244	259	273
04	287	302	316	330	344	359	373	387	401	416
05	430	444	458	473	487	501	515	530	544	558
06	572	586	601	615	629	643	657	671	686	700
07	714	728	742	756	770	785	799	813	827	841
08	855	869	883	897	911	926	940	954	968	982
09	48 996	*010	*024	*038	*052	*066	*080	*094	*108	*122
310	49 136	150	164	178	192	206	220	234	248	262
11	276	290	304	318	332	346	360	374	388	402
12	415	429	443	457	471	485	499	513	527	541
13	554	568	582	596	610	624	638	651	665	679
14	693	707	721	734	748	762	776	790	803	817
15	831	845	859	872	886	900	914	927	941	955
16	49 969	982	996	*010	*024	*037	*051	*065	*079	*092
17	50 106	120	133	147	161	174	188	202	215	229
18	243	256	270	284	297	311	325	338	352	365
19	379	393	406	420	433	447	461	474	488	501
320	515	529	542	556	569	583	596	610	623	637
21	651	664	678	691	705	718	732	745	759	772
22	786	799	813	826	840	853	866	880	893	907
23	50 920	934	947	961	974	987	*001	*014	*028	*041
24	51 055	068	081	095	108	121	135	148	162	175
25	188	202	215	228	242	255	268	282	295	308
26	322	335	348	362	375	388	402	415	428	441
27	455	468	481	495	508	521	534	548	561	574
28	587	601	614	627	640	654	667	680	693	706
29	720	733	746	759	772	786	799	812	825	838
330	851	865	878	891	904	917	930	943	957	970
31	51 983	996	*009	*022	*035	*048	*061	*075	*088	*101
32	52 114	127	140	153	166	179	192	205	218	231
33	244	257	270	284	297	310	323	336	349	362
34	375	388	401	414	427	440	453	466	479	492
35	504	517	530	543	556	569	582	595	608	621
36	634	647	660	673	686	699	711	724	737	750
37	763	776	789	802	815	827	840	853	866	879
38	52 892	905	917	930	943	956	969	982	994	*007
39	53 020	033	046	058	071	084	097	110	122	135
340	148	161	173	186	199	212	224	237	250	263
41	275	288	301	314	326	339	352	364	377	390
42	403	415	428	441	453	466	479	491	504	517
43	529	542	555	567	580	593	605	618	631	643
44	656	668	681	694	706	719	732	744	757	769
45	782	794	807	820	832	845	857	870	882	895
46	53 908	920	933	945	958	970	983	995	*008	*020
47	54 033	045	058	070	083	095	108	120	133	145
48	158	170	183	195	208	220	233	245	258	270
49	283	295	307	320	332	345	357	370	382	394
350	54 407	419	432	444	456	469	481	494	506	518
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

	15	14	13	12
1	1.5	1.4	1.3	1.2
2	3.0	2.8	2.6	2.4
3	4.5	4.2	3.9	3.6
4	6.0	5.6	5.2	4.8
5	7.5	7.0	6.5	6.0
6	9.0	8.4	7.8	7.2
7	10.5	9.8	9.1	8.4
8	12.0	11.2	10.4	9.6
9	13.5	12.6	11.7	10.8

Prop. Parts

TABLE V—(*Continued*)
FIVE-PLACE LOGARITHMS: 350–400

N	0	1	2	3	4	5	6	7	8	9
350	54 407	419	432	444	456	469	481	494	506	518
51	531	543	555	568	580	593	605	617	630	642
52	654	667	679	691	704	716	728	741	753	765
53	777	790	802	814	827	839	851	864	876	888
54	54 900	913	925	937	949	962	974	986	998	*011
55	55 023	035	047	060	072	084	096	108	121	133
56	145	157	169	182	194	206	218	230	242	255
57	267	279	291	303	315	328	340	352	364	376
58	388	400	413	425	437	449	461	473	485	497
59	509	522	534	546	558	570	582	594	606	618
360	630	642	654	666	678	691	703	715	727	739
61	751	763	775	787	799	811	823	835	847	859
62	871	883	895	907	919	931	943	955	967	979
63	55 991	*003	*015	*027	*038	*050	*062	*074	*086	*098
64	56 110	122	134	146	158	170	182	194	205	217
65	229	241	253	265	277	289	301	312	324	336
66	348	360	372	384	396	407	419	431	443	455
67	467	478	490	502	514	526	538	549	561	573
68	585	597	608	620	632	644	656	667	679	691
69	703	714	726	738	750	761	773	785	797	808
370	820	832	844	855	867	879	891	902	914	926
71	56 937	949	961	972	984	996	*008	*019	*031	*043
72	57 054	066	078	089	101	113	124	136	148	159
73	171	183	194	206	217	229	241	252	264	276
74	287	299	310	322	334	345	357	368	380	392
75	403	415	426	438	449	461	473	484	496	507
76	519	530	542	553	565	576	588	600	611	623
77	634	646	657	669	680	692	703	715	726	738
78	749	761	772	784	795	807	818	830	841	852
79	864	875	887	898	910	921	933	944	955	967
380	57 978	990	*001	*013	*024	*035	*047	*058	*070	*081
81	58 092	104	115	127	138	149	161	172	184	195
82	206	218	229	240	252	263	274	286	297	309
83	320	331	343	354	365	377	388	399	410	422
84	433	444	456	467	478	490	501	512	524	535
85	546	557	569	580	591	602	614	625	636	647
86	659	670	681	692	704	715	726	737	749	760
87	771	782	794	805	816	827	838	850	861	872
88	883	894	906	917	928	939	950	961	973	984
89	58 995	*006	*017	*028	*040	*051	*062	*073	*084	*095
390	59 106	118	129	140	151	162	173	184	195	207
91	218	229	240	251	262	273	284	295	306	318
92	329	340	351	362	373	384	395	406	417	428
93	439	450	461	472	483	494	506	517	528	539
94	550	561	572	583	594	605	616	627	638	649
95	660	671	682	693	704	715	726	737	748	759
96	770	780	791	802	813	824	835	846	857	868
97	879	890	901	912	923	934	945	956	966	977
98	59 988	999	*010	*021	*032	*043	*054	*065	*076	*086
99	60 097	108	119	130	141	152	163	173	184	195
400	60 206	217	228	239	249	260	271	282	293	304
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

	13	12	11	10
1	1.3	1.2	1.1	1.0
2	2.6	2.4	2.2	2.0
3	3.9	3.6	3.3	3.0
4	5.2	4.8	4.4	4.0
5	6.5	6.0	5.5	5.0
6	7.8	7.2	6.6	6.0
7	9.1	8.4	7.7	7.0
8	10.4	9.6	8.8	8.0
9	11.7	10.8	9.9	9.0

TABLE V—(*Continued*)

FIVE-PLACE LOGARITHMS. 400–450

N	0	1	2	3	4	5	6	7	8	9
400	60 206	217	228	239	249	260	271	282	293	304
01	314	325	336	347	358	369	379	390	401	412
02	423	433	444	455	466	477	487	498	509	520
03	531	541	552	563	574	584	595	606	617	627
04	638	649	660	670	681	692	703	713	724	735
05	746	756	767	778	788	799	810	821	831	842
06	853	863	874	885	895	906	917	927	938	949
07	60 959	970	981	991	*002	*013	*023	*034	*045	*055
08	61 066	077	087	098	109	119	130	140	151	162
09	172	183	194	204	215	225	236	247	257	268
410	278	289	300	310	321	331	342	352	363	374
11	384	395	405	416	426	437	448	458	469	479
12	490	500	511	521	532	542	553	563	574	584
13	595	606	616	627	637	648	658	669	679	690
14	700	711	721	731	742	752	763	773	784	794
15	805	815	826	836	847	857	868	878	888	899
16	61 909	920	930	941	951	962	972	982	993	*003
17	62 014	024	034	045	055	066	076	086	097	107
18	118	128	138	149	159	170	180	190	201	211
19	221	232	242	252	263	273	284	294	304	315
420	325	335	346	356	366	377	387	397	408	418
21	428	439	449	459	469	480	490	500	511	521
22	531	542	552	562	572	583	593	603	613	624
23	634	644	655	665	675	685	696	706	716	726
24	737	747	757	767	778	788	798	808	818	829
25	839	849	859	870	880	890	900	910	921	931
26	62 941	951	961	972	982	992	*002	*012	*022	*033
27	63 043	053	063	073	083	094	104	114	124	134
28	144	155	165	175	185	195	205	215	225	236
29	246	256	266	276	286	296	306	317	327	337
430	347	357	367	377	387	397	407	417	428	438
31	448	458	468	478	488	498	508	518	528	538
32	548	558	568	579	589	599	609	619	629	639
33	649	659	669	679	689	699	709	719	729	739
34	749	759	769	779	789	799	809	819	829	839
35	849	859	869	879	889	899	909	919	929	939
36	63 949	959	969	979	988	998	*008	*018	*028	*038
37	64 048	058	068	078	088	098	108	118	128	137
38	147	157	167	177	187	197	207	217	227	237
39	246	256	266	276	286	296	306	316	326	335
440	345	355	365	375	385	395	404	414	424	434
41	444	454	464	473	483	493	503	513	523	532
42	542	552	562	572	582	591	601	611	621	631
43	640	650	660	670	680	689	699	709	719	729
44	738	748	758	768	777	787	797	807	816	826
45	836	846	856	865	875	885	895	904	914	924
46	64 933	943	953	963	972	982	992	*002	*011	*021
47	65 031	040	050	060	070	079	089	099	108	118
48	128	137	147	157	167	176	186	196	205	215
49	225	234	244	254	263	273	283	292	302	312
450	65 321	331	341	350	360	369	379	389	398	408
N	0	1	2	3	4	5	6	7	8	9

Prop. Parts

	11
1	1.1
2	2.2
3	3 3
4	4 4
5	5 5
6	6.6
7	7.7
8	8.8
9	9.9

	10
1	1 0
2	2 0
3	3.0
4	4 0
5	5 0
6	6 0
7	7 0
8	8 0
9	9.0

	9
1	0 9
2	1 8
3	2.7
4	3.6
5	4 5
6	5.4
7	6 3
8	7.2
9	8.1

Prop. Parts

TABLE V—(*Continued*)
FIVE-PLACE LOGARITHMS: 450–500

N	0	1	2	3	4	5	6	7	8	9
450	65 321	331	341	350	360	369	379	389	398	408
51	418	427	437	447	456	466	475	485	495	504
52	514	523	533	543	552	562	571	581	591	600
53	610	619	629	639	648	658	667	677	686	696
54	706	715	725	734	744	753	763	772	782	792
55	801	811	820	830	839	849	858	868	877	887
56	896	906	916	925	935	944	954	963	973	982
57	65 992	*001	*011	*020	*030	*039	*049	*058	*068	*077
58	66 087	096	106	115	124	134	143	153	162	172
59	181	191	200	210	219	229	238	247	257	266
460	276	285	295	304	314	323	332	342	351	361
61	370	380	389	398	408	417	427	436	445	455
62	464	474	483	492	502	511	521	530	539	549
63	558	567	577	586	596	605	614	624	633	642
64	652	661	671	680	689	699	708	717	727	736
65	745	755	764	773	783	792	801	811	820	829
66	839	848	857	867	876	885	894	904	913	922
67	66 932	941	950	960	969	978	987	997	*006	*015
68	67 025	034	043	052	062	071	080	089	099	108
69	117	127	136	145	154	164	173	182	191	201
470	210	219	228	237	247	256	265	274	284	293
71	302	311	321	330	339	348	357	367	376	385
72	394	403	413	422	431	440	449	459	468	477
73	486	495	504	514	523	532	541	550	560	569
74	578	587	596	605	614	624	633	642	651	660
75	669	679	688	697	706	715	724	733	742	752
76	761	770	779	788	797	806	815	825	834	843
77	852	861	870	879	888	897	906	916	925	934
78	67 943	952	961	970	979	988	997	*006	*015	*024
79	68 034	043	052	061	070	079	088	097	106	115
480	124	133	142	151	160	169	178	187	196	205
81	215	224	233	242	251	260	269	278	287	296
82	305	314	323	332	341	350	359	368	377	386
83	395	404	413	422	431	440	449	458	467	476
84	485	494	502	511	520	529	538	547	556	565
85	574	583	592	601	610	619	628	637	646	655
86	664	673	681	690	699	708	717	726	735	744
87	753	762	771	780	789	797	806	815	824	833
88	842	851	860	869	878	886	895	904	913	922
89	68 931	940	949	958	966	975	984	993	*002	*011
490	69 020	028	037	046	055	064	073	082	090	099
91	108	117	126	135	144	152	161	170	179	188
92	197	205	214	223	232	241	249	258	267	276
93	285	294	302	311	320	329	338	346	355	364
94	373	381	390	399	408	417	425	434	443	452
95	461	469	478	487	496	504	513	522	531	539
96	548	557	566	574	583	592	601	609	618	627
97	636	644	653	662	671	679	688	697	705	714
98	723	732	740	749	758	767	775	784	793	801
99	810	819	827	836	845	854	862	871	880	888
500	69 897	906	914	923	932	940	949	958	966	975
N	0	1	2	3	4	5	6	7	8	9

Prop. Parts

	10
1	1.0
2	2.0
3	3.0
4	4.0
5	5.0
6	6.0
7	7.0
8	8.0
9	9.0

	9
1	0.9
2	1.8
3	2.7
4	3.6
5	4.5
6	5.4
7	6.3
8	7.2
9	8.1

	8
1	0.8
2	1.6
3	2.4
4	3.2
5	4.0
6	4.8
7	5.6
8	6.4
9	7.2

Prop. Parts

TABLE V—(*Continued*)

FIVE-PLACE LOGARITHMS: 500–550

N	0	1	2	3	4	5	6	7	8	9	Prop. Parts
500	69 897	906	914	923	932	940	949	958	966	975	
01	69 984	992	*001	*010	*018	*027	*036	*044	*053	*062	
02	70 070	079	088	096	105	114	122	131	140	148	
03	157	165	174	183	191	200	209	217	226	234	
04	243	252	260	269	278	286	295	303	312	321	
05	329	338	346	355	364	372	381	389	398	406	
06	415	424	432	441	449	458	467	475	484	492	
07	501	509	518	526	535	544	552	561	569	578	
08	586	595	603	612	621	629	638	646	655	663	
09	672	680	689	697	706	714	723	731	740	749	
510	757	766	774	783	791	800	808	817	825	834	
11	842	851	859	868	876	885	893	902	910	919	
12	70 927	935	944	952	961	969	978	986	995	*003	
13	71 012	020	029	037	046	054	063	071	079	088	
14	096	105	113	122	130	139	147	155	164	172	
15	181	189	198	206	214	223	231	240	248	257	
16	265	273	282	290	299	307	315	324	332	341	
17	349	357	366	374	383	391	399	408	416	425	
18	433	441	450	458	466	475	483	492	500	508	
19	517	525	533	542	550	559	567	575	584	592	
520	600	609	617	625	634	642	650	659	667	675	
21	684	692	700	709	717	725	734	742	750	759	
22	767	775	784	792	800	809	817	825	834	842	
23	850	858	867	875	883	892	900	908	917	925	
24	71 933	941	950	958	966	975	983	991	999	*008	
25	72 016	024	032	041	049	057	066	074	082	090	
26	099	107	115	123	132	140	148	156	165	173	
27	181	189	198	206	214	222	230	239	247	255	
28	263	272	280	288	296	304	313	321	329	337	
29	346	354	362	370	378	387	395	403	411	419	
530	428	436	444	452	460	469	477	485	493	501	
31	509	518	526	534	542	550	558	567	575	583	
32	591	599	607	616	624	632	640	648	656	665	
33	673	681	689	697	705	713	722	730	738	746	
34	754	762	770	779	787	795	803	811	819	827	
35	835	843	852	860	868	876	884	892	900	908	
36	916	925	933	941	949	957	965	973	981	989	
37	72 997	*006	*014	*022	*030	*038	*046	*054	*062	*070	
38	73 078	086	094	102	111	119	127	135	143	151	
39	159	167	175	183	191	199	207	215	223	231	
540	239	247	255	263	272	280	288	296	304	312	
41	320	328	336	344	352	360	368	376	384	392	
42	400	408	416	424	432	440	448	456	464	472	
43	480	488	496	504	512	520	528	536	544	552	
44	560	568	576	584	592	600	608	616	624	632	
45	640	648	656	664	672	679	687	695	703	711	
46	719	727	735	743	751	759	767	775	783	791	
47	799	807	815	823	830	838	846	854	862	870	
48	878	886	894	902	910	918	926	933	941	949	
49	73 957	965	973	981	989	997	*005	*013	*020	*028	
550	74 036	044	052	060	068	076	084	092	099	107	
N	0	1	2	3	4	5	6	7	8	9	Prop. Parts

Prop. Parts:

	9
1	0.9
2	1.8
3	2.7
4	3.6
5	4.5
6	5.4
7	6.3
8	7.2
9	8.1

	8
1	0.8
2	1.6
3	2.4
4	3.2
5	4.0
6	4.8
7	5.6
8	6.4
9	7.2

	7
1	0.7
2	1.4
3	2.1
4	2.8
5	3.5
6	4.2
7	4.9
8	5.6
9	6.3

TABLE V—(*Continued*)

FIVE-PLACE LOGARITHMS: 550–600

N	0	1	2	3	4	5	6	7	8	9
550	74 036	044	052	060	068	076	084	092	099	107
51	115	123	131	139	147	155	162	170	178	186
52	194	202	210	218	225	233	241	249	257	265
53	273	280	288	296	304	312	320	327	335	343
54	351	359	367	374	382	390	398	406	414	421
55	429	437	445	453	461	468	476	484	492	500
56	507	515	523	531	539	547	554	562	570	578
57	586	593	601	609	617	624	632	640	648	656
58	663	671	679	687	695	702	710	718	726	733
59	741	749	757	764	772	780	788	796	803	811
560	819	827	834	842	850	858	865	873	881	889
61	896	904	912	920	927	935	943	950	958	966
62	74 974	981	989	997	*005	*012	*020	*028	*035	*043
63	75 051	059	066	074	082	089	097	105	113	120
64	128	136	143	151	159	166	174	182	189	197
65	205	213	220	228	236	243	251	259	266	274
66	282	289	297	305	312	320	328	335	343	351
67	358	366	374	381	389	397	404	412	420	427
68	435	442	450	458	465	473	481	488	496	504
69	511	519	526	534	542	549	557	565	572	580
570	587	595	603	610	618	626	633	641	648	656
71	664	671	679	686	694	702	709	717	724	732
72	740	747	755	762	770	778	785	793	800	808
73	815	823	831	838	846	853	861	868	876	884
74	891	899	906	914	921	929	937	944	952	959
75	75 967	974	982	989	997	*005	*012	*020	*027	*035
76	76 042	050	057	065	072	080	087	095	103	110
77	118	125	133	140	148	155	163	170	178	185
78	193	200	208	215	223	230	238	245	253	260
79	268	275	283	290	298	305	313	320	328	335
580	343	350	358	365	373	380	388	395	403	410
81	418	425	433	440	448	455	462	470	477	485
82	492	500	507	515	522	530	537	545	552	559
83	567	574	582	589	597	604	612	619	626	634
84	641	649	656	664	671	678	686	693	701	708
85	716	723	730	738	745	753	760	768	775	782
86	790	797	805	812	819	827	834	842	849	856
87	864	871	879	886	893	901	908	916	923	930
88	76 938	945	953	960	967	975	982	989	997	*004
89	77 012	019	026	034	041	048	056	063	070	078
590	085	093	100	107	115	122	129	137	144	151
91	159	166	173	181	188	195	203	210	217	225
92	232	240	247	254	262	269	276	283	291	298
93	305	313	320	327	335	342	349	357	364	371
94	379	386	393	401	408	415	422	430	437	444
95	452	459	466	474	481	488	495	503	510	517
96	525	532	539	546	554	561	568	576	583	590
97	597	605	612	619	627	634	641	648	656	663
98	670	677	685	692	699	706	714	721	728	735
99	743	750	757	764	772	779	786	793	801	808
600	77 815	822	830	837	844	851	859	866	873	880
N	0	1	2	3	4	5	6	7	8	9

Prop. Parts

	8
1	0.8
2	1 6
3	2.4
4	3 2
5	4 0
6	4 8
7	5 6
8	6 4
9	7.2

	7
1	0.7
2	1.4
3	2.1
4	2 8
5	3 5
6	4.2
7	4 9
8	5 6
9	6.3

TABLE V—(*Continued*)
FIVE-PLACE LOGARITHMS: 600–650

N	0	1	2	3	4	5	6	7	8	9
600	77 815	822	830	837	844	851	859	866	873	880
01	887	895	902	909	916	924	931	938	945	952
02	77 960	967	974	981	988	996	*003	010	*017	*025
03	78 032	039	046	053	061	068	075	082	089	097
04	104	111	118	125	132	140	147	154	161	168
05	176	183	190	197	204	211	219	226	233	240
06	247	254	262	269	276	283	290	297	305	312
07	319	326	333	340	347	355	362	369	376	383
08	390	398	405	412	419	426	433	440	447	455
09	462	469	476	483	490	497	504	512	519	526
610	533	540	547	554	561	569	576	583	590	597
11	604	611	618	625	633	640	647	654	661	668
12	675	682	689	696	704	711	718	725	732	739
13	746	753	760	767	774	781	789	796	803	810
14	817	824	831	838	845	852	859	866	873	880
15	888	895	902	909	916	923	930	937	944	951
16	78 958	965	972	979	986	993	*000	*007	*014	*021
17	79 029	036	043	050	057	064	071	078	085	092
18	099	106	113	120	127	134	141	148	155	162
19	169	176	183	190	197	204	211	218	225	232
620	239	246	253	260	267	274	281	288	295	302
21	309	316	323	330	337	344	351	358	365	372
22	379	386	393	400	407	414	421	428	435	442
23	449	456	463	470	477	484	491	498	505	511
24	518	525	532	539	546	553	560	567	574	581
25	588	595	602	609	616	623	630	637	644	650
26	657	664	671	678	685	692	699	706	713	720
27	727	734	741	748	754	761	768	775	782	789
28	796	803	810	817	824	831	837	844	851	858
29	865	872	879	886	893	900	906	913	920	927
630	79 934	941	948	955	962	969	975	982	989	996
31	80 003	010	017	024	030	037	044	051	058	065
32	072	079	085	092	099	106	113	120	127	134
33	140	147	154	161	168	175	182	188	195	202
34	209	216	223	229	236	243	250	257	264	271
35	277	284	291	298	305	312	318	325	332	339
36	346	353	359	366	373	380	387	393	400	407
37	414	421	428	434	441	448	455	462	468	475
38	482	489	496	502	509	516	523	530	536	543
39	550	557	564	570	577	584	591	598	604	611
640	618	625	632	638	645	652	659	665	672	679
41	686	693	699	706	713	720	726	733	740	747
42	754	760	767	774	781	787	794	801	808	814
43	821	828	835	841	848	855	862	868	875	882
44	889	895	902	909	916	922	929	936	943	949
45	80 956	963	969	976	983	990	996	*003	*010	*017
46	81 023	030	037	043	050	057	064	070	077	084
47	090	097	104	111	117	124	131	137	144	151
48	158	164	171	178	184	191	198	204	211	218
49	224	231	238	245	251	258	265	271	278	285
650	81 291	298	305	311	318	325	331	338	345	351
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

	8	7	6
1	0.8	0.7	0.6
2	1.6	1.4	1.2
3	2.4	2.1	1.8
4	3.2	2.8	2.4
5	4.0	3.5	3.0
6	4.8	4.2	3.6
7	5.6	4.9	4.2
8	6.4	5.6	4.8
9	7.2	6.3	5.4

TABLE V—(*Continued*)
FIVE-PLACE LOGARITHMS: 650–700

Prop. Parts

	7
1	0.7
2	1.4
3	2.1
4	2.8
5	3.5
6	4.2
7	4.9
8	5.6
9	6.3

	6
1	0.6
2	1.2
3	1.8
4	2.4
5	3.0
6	3.6
7	4.2
8	4.8
9	5.4

N	0	1	2	3	4	5	6	7	8	9
650	81 291	298	305	311	318	325	331	338	345	351
51	358	365	371	378	385	391	398	405	411	418
52	425	431	438	445	451	458	465	471	478	485
53	491	498	505	511	518	525	531	538	544	551
54	558	564	571	578	584	591	598	604	611	617
55	624	631	637	644	651	657	664	671	677	684
56	690	697	704	710	717	723	730	737	743	750
57	757	763	770	776	783	790	796	803	809	816
58	823	829	836	842	849	856	862	869	875	882
59	889	895	902	908	915	921	928	935	941	948
660	81 954	961	968	974	981	987	994	*000	*007	*014
61	82 020	027	033	040	046	053	060	066	073	079
62	086	092	099	105	112	119	125	132	138	145
63	151	158	164	171	178	184	191	197	204	210
64	217	223	230	236	243	249	256	263	269	276
65	282	289	295	302	308	315	321	328	334	341
66	347	354	360	367	373	380	387	393	400	406
67	413	419	426	432	439	445	452	458	465	471
68	478	484	491	497	504	510	517	523	530	536
69	543	549	556	562	569	575	582	588	595	601
670	607	614	620	627	633	640	646	653	659	666
71	672	679	685	692	698	705	711	718	724	730
72	737	743	750	756	763	769	776	782	789	795
73	802	808	814	821	827	834	840	847	853	860
74	866	872	879	885	892	898	905	911	918	924
75	930	937	943	950	956	963	969	975	982	988
76	82 995	*001	*008	*014	*020	*027	*033	*040	*046	*052
77	83 059	065	072	078	085	091	097	104	110	117
78	123	129	136	142	149	155	161	168	174	181
79	187	193	200	206	213	219	225	232	238	245
680	251	257	264	270	276	283	289	296	302	308
81	315	321	327	334	340	347	353	359	366	372
82	378	385	391	398	404	410	417	423	429	436
83	442	448	455	461	467	474	480	487	493	499
84	506	512	518	525	531	537	544	550	556	563
85	569	575	582	588	594	601	607	613	620	626
86	632	639	645	651	658	664	670	677	683	689
87	696	702	708	715	721	727	734	740	746	753
88	759	765	771	778	784	790	797	803	809	816
89	822	828	835	841	847	853	860	866	872	879
690	885	891	897	904	910	916	923	929	935	942
91	83 948	954	960	967	973	979	985	992	998	*004
92	84 011	017	023	029	036	042	048	055	061	067
93	073	080	086	092	098	105	111	117	123	130
94	136	142	148	155	161	167	173	180	186	192
95	198	205	211	217	223	230	236	242	248	255
96	261	267	273	280	286	292	298	305	311	317
97	323	330	336	342	348	354	361	367	373	379
98	386	392	398	404	410	417	423	429	435	442
99	448	454	460	466	473	479	485	491	497	504
700	84 510	516	522	528	535	541	547	553	559	566
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

TABLE V—(*Continued*)

FIVE-PLACE LOGARITHMS. 700–750

N	0	1	2	3	4	5	6	7	8	9	Prop. Parts
700	84 510	516	522	528	535	541	547	553	559	566	
01	572	578	584	590	597	603	609	615	621	628	
02	634	640	646	652	658	665	671	677	683	689	
03	696	702	708	714	720	726	733	739	745	751	
04	757	763	770	776	782	788	794	800	807	813	
05	819	825	831	837	844	850	856	862	868	874	
06	880	887	893	899	905	911	917	924	930	936	
07	84 942	948	954	960	967	973	979	985	991	997	
08	85 003	009	016	022	028	034	040	046	052	058	
09	065	071	077	083	089	095	101	107	114	120	
710	126	132	138	144	150	156	163	169	175	181	
11	187	193	199	205	211	217	224	230	236	242	
12	248	254	260	266	272	278	285	291	297	303	
13	309	315	321	327	333	339	345	352	358	364	
14	370	376	382	388	394	400	406	412	418	425	
15	431	437	443	449	455	461	467	473	479	485	
16	491	497	503	509	516	522	528	534	540	546	
17	552	558	564	570	576	582	588	594	600	606	
18	612	618	625	631	637	643	649	655	661	667	
19	673	679	685	691	697	703	709	715	721	727	
720	733	739	745	751	757	763	769	775	781	788	
21	794	800	806	812	818	824	830	836	842	848	
22	854	860	866	872	878	884	890	896	902	908	
23	914	920	926	932	938	944	950	956	962	968	
24	85 974	980	986	992	998	*004	*010	*016	*022	*028	
25	86 034	040	046	052	058	064	070	076	082	088	
26	094	100	106	112	118	124	130	136	141	147	
27	153	159	165	171	177	183	189	195	201	207	
28	213	219	225	231	237	243	249	255	261	267	
29	273	279	285	291	297	303	308	314	320	326	
730	332	338	344	350	356	362	368	374	380	386	
31	392	398	404	410	415	421	427	433	439	445	
32	451	457	463	469	475	481	487	493	499	504	
33	510	516	522	528	534	540	546	552	558	564	
34	570	576	581	587	593	599	605	611	617	623	
35	629	635	641	646	652	658	664	670	676	682	
36	688	694	700	705	711	717	723	729	735	741	
37	747	753	759	764	770	776	782	788	794	800	
38	806	812	817	823	829	835	841	847	853	859	
39	864	870	876	882	888	894	900	906	911	917	
740	923	929	935	941	947	953	958	964	970	976	
41	86 982	988	994	999	*005	*011	*017	*023	*029	*035	
42	87 040	046	052	058	064	070	075	081	087	093	
43	099	105	111	116	122	128	134	140	146	151	
44	157	163	169	175	181	186	192	198	204	210	
45	216	221	227	233	239	245	251	256	262	268	
46	274	280	286	291	297	303	309	315	320	326	
47	332	338	344	349	355	361	367	373	379	384	
48	390	396	402	408	413	419	425	431	437	442	
49	448	454	460	466	471	477	483	489	495	500	
750	87 506	512	518	523	529	535	541	547	552	558	
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**Prop. Parts**

Prop. Parts:

	7
1	0.7
2	1.4
3	2.1
4	2.8
5	3.5
6	4.2
7	4.9
8	5.6
9	6.3

	6
1	0.6
2	1.2
3	1.8
4	2.4
5	3.0
6	3.6
7	4.2
8	4.8
9	5.4

	5
1	0.5
2	1.0
3	1.5
4	2.0
5	2.5
6	3.0
7	3.5
8	4.0
9	4.5

TABLE V—(*Continued*)
FIVE-PLACE LOGARITHMS: 750–800

N	0	1	2	3	4	5	6	7	8	9
750	87 506	512	518	523	529	535	541	547	552	558
51	564	570	576	581	587	593	599	604	610	616
52	622	628	633	639	645	651	656	662	668	674
53	679	685	691	697	703	708	714	720	726	731
54	737	743	749	754	760	766	772	777	783	789
55	795	800	806	812	818	823	829	835	841	846
56	852	858	864	869	875	881	887	892	898	904
57	910	915	921	927	933	938	944	950	955	961
58	87 967	973	978	984	990	996	*001	*007	*013	*018
59	88 024	030	036	041	047	053	058	064	070	076
760	081	087	093	098	104	110	116	121	127	133
61	138	144	150	156	161	167	173	178	184	190
62	195	201	207	213	218	224	230	235	241	247
63	252	258	264	270	275	281	287	292	298	304
64	309	315	321	326	332	338	343	349	355	360
65	366	372	377	383	389	395	400	406	412	417
66	423	429	434	440	446	451	457	463	468	474
67	480	485	491	497	502	508	513	519	525	530
68	536	542	547	553	559	564	570	576	581	587
69	593	598	604	610	615	621	627	632	638	643
770	649	655	660	666	672	677	683	689	694	700
71	705	711	717	722	728	734	739	745	750	756
72	762	767	773	779	784	790	795	801	807	812
73	818	824	829	835	840	846	852	857	863	868
74	874	880	885	891	897	902	908	913	919	925
75	930	936	941	947	953	958	964	969	975	981
76	88 986	992	997	*003	*009	*014	*020	*025	*031	*037
77	89 042	048	053	059	064	070	076	081	087	092
78	098	104	109	115	120	126	131	137	143	148
79	154	159	165	170	176	182	187	193	198	204
780	209	215	221	226	232	237	243	248	254	260
81	265	271	276	282	287	293	298	304	310	315
82	321	326	332	337	343	348	354	360	365	371
83	376	382	387	393	398	404	409	415	421	426
84	432	437	443	448	454	459	465	470	476	481
85	487	492	498	504	509	515	520	526	531	537
86	542	548	553	559	564	570	575	581	586	592
87	597	603	609	614	620	625	631	636	642	647
88	653	658	664	669	675	680	686	691	697	702
89	708	713	719	724	730	735	741	746	752	757
790	763	768	774	779	785	790	796	801	807	812
91	818	823	829	834	840	845	851	856	862	867
92	873	878	883	889	894	900	905	911	916	922
93	927	933	938	944	949	955	960	966	971	977
94	89 982	988	993	998	*004	*009	*015	*020	*026	*031
95	90 037	042	048	053	059	064	069	075	080	086
96	091	097	102	108	113	119	124	129	135	140
97	146	151	157	162	168	173	179	184	189	195
98	200	206	211	217	222	227	233	238	244	249
99	255	260	266	271	276	282	287	293	298	304
800	90 309	314	320	325	331	336	342	347	352	358
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

	6
1	0.6
2	1.2
3	1.8
4	2.4
5	3.0
6	3.6
7	4.2
8	4.8
9	5.4

	5
1	0.5
2	1.0
3	1.5
4	2.0
5	2.5
6	3.0
7	3.5
8	4.0
9	4.5

Prop. Parts

TABLE V—(*Continued*)

FIVE-PLACE LOGARITHMS: 800–850

N	0	1	2	3	4	5	6	7	8	9
800	90 309	314	320	325	331	336	342	347	352	358
01	363	369	374	380	385	390	396	401	407	412
02	417	423	428	434	439	445	450	455	461	466
03	472	477	482	488	493	499	504	509	515	520
04	526	531	536	542	547	553	558	563	569	574
05	580	585	590	596	601	607	612	617	623	628
06	634	639	644	650	655	660	666	671	677	682
07	687	693	698	703	709	714	720	725	730	736
08	741	747	752	757	763	768	773	779	784	789
09	795	800	806	811	816	822	827	832	838	843
810	849	854	859	865	870	875	881	886	891	897
11	902	907	913	918	924	929	934	940	945	950
12	90 956	961	966	972	977	982	988	993	998	*004
13	91 009	014	020	025	030	036	041	046	052	057
14	062	068	073	078	084	089	094	100	105	110
15	116	121	126	132	137	142	148	153	158	164
16	169	174	180	185	190	196	201	206	212	217
17	222	228	233	238	243	249	254	259	265	270
18	275	281	286	291	297	302	307	312	318	323
19	328	334	339	344	350	355	360	365	371	376
820	381	387	392	397	403	408	413	418	424	429
21	434	440	445	450	455	461	466	471	477	482
22	487	492	498	503	508	514	519	524	529	535
23	540	545	551	556	561	566	572	577	582	587
24	593	598	603	609	614	619	624	630	635	640
25	645	651	656	661	666	672	677	682	687	693
26	698	703	709	714	719	724	730	735	740	745
27	751	756	761	766	772	777	782	787	793	798
28	803	808	814	819	824	829	834	840	845	850
29	855	861	866	871	876	882	887	892	897	903
830	908	913	918	924	929	934	939	944	950	955
31	91 960	965	971	976	981	986	991	997	*002	*007
32	92 012	018	023	028	033	038	044	049	054	059
33	065	070	075	080	085	091	096	101	106	111
34	117	122	127	132	137	143	148	153	158	163
35	169	174	179	184	189	195	200	205	210	215
36	221	226	231	236	241	247	252	257	262	267
37	273	278	283	288	293	298	304	309	314	319
38	324	330	335	340	345	350	355	361	366	371
39	376	381	387	392	397	402	407	412	418	423
840	428	433	438	443	449	454	459	464	469	474
41	480	485	490	495	500	505	511	516	521	526
42	531	536	542	547	552	557	562	567	572	578
43	583	588	593	598	603	609	614	619	624	629
44	634	639	645	650	655	660	665	670	675	681
45	686	691	696	701	706	711	716	722	727	732
46	737	742	747	752	758	763	768	773	778	783
47	788	793	799	804	809	814	819	824	829	834
48	840	845	850	855	860	865	870	875	881	886
49	891	896	901	906	911	916	921	927	932	937
850	92 942	947	952	957	962	967	973	978	983	988
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

	6
1	0 6
2	1 2
3	1 8
4	2 4
5	3 0
6	3 6
7	4 2
8	4 8
9	5.4

	5
1	0 5
2	1.0
3	1 5
4	2 0
5	2 0
6	3.0
7	3.5
8	4.0
9	4.5

Prop. Parts

TABLE V—(*Continued*)
FIVE-PLACE LOGARITHMS: 850–900

N	0	1	2	3	4	5	6	7	8	9
850	92 942	947	952	957	962	967	973	978	983	988
51	92 993	998	*003	*008	*013	*018	*024	*029	*034	*039
52	93 044	049	054	059	064	069	075	080	085	090
53	095	100	105	110	115	120	125	131	136	141
54	146	151	156	161	166	171	176	181	186	192
55	197	202	207	212	217	222	227	232	237	242
56	247	252	258	263	268	273	278	283	288	293
57	298	303	308	313	318	323	328	334	339	344
58	349	354	359	364	369	374	379	384	389	394
59	399	404	409	414	420	425	430	435	440	445
860	450	455	460	465	470	475	480	485	490	495
61	500	505	510	515	520	526	531	536	541	546
62	551	556	561	566	571	576	581	586	591	596
63	601	606	611	616	621	626	631	636	641	646
64	651	656	661	666	671	676	682	687	692	697
65	702	707	712	717	722	727	732	737	742	747
66	752	757	762	767	772	777	782	787	792	797
67	802	807	812	817	822	827	832	837	842	847
68	852	857	862	867	872	877	882	887	892	897
69	902	907	912	917	922	927	932	937	942	947
870	93 952	957	962	967	972	977	982	987	992	997
71	94 002	007	012	017	022	027	032	037	042	047
72	052	057	062	067	072	077	082	086	091	096
73	101	106	111	116	121	126	131	136	141	146
74	151	156	161	166	171	176	181	186	191	196
75	201	206	211	216	221	226	231	236	240	245
76	250	255	260	265	270	275	280	285	290	295
77	300	305	310	315	320	325	330	335	340	345
78	349	354	359	364	369	374	379	384	389	394
79	399	404	409	414	419	424	429	433	438	443
880	448	453	458	463	468	473	478	483	488	493
81	498	503	507	512	517	522	527	532	537	542
82	547	552	557	562	567	571	576	581	586	591
83	596	601	606	611	616	621	626	630	635	640
84	645	650	655	660	665	670	675	680	685	689
85	694	699	704	709	714	719	724	729	734	738
86	743	748	753	758	763	768	773	778	783	787
87	792	797	802	807	812	817	822	827	832	836
88	841	846	851	856	861	866	871	876	880	885
89	890	895	900	905	910	915	919	924	929	934
890	939	944	949	954	959	963	968	973	978	983
91	94 988	993	998	*002	*007	*012	*017	*022	*027	*032
92	95 036	041	046	051	056	061	066	071	075	080
93	085	090	095	100	105	109	114	119	124	129
94	134	139	143	148	153	158	163	168	173	177
95	182	187	192	197	202	207	211	216	221	226
96	231	236	240	245	250	255	260	265	270	274
97	279	284	289	294	299	303	308	313	318	323
98	328	332	337	342	347	352	357	361	366	371
99	376	381	386	390	395	400	405	410	415	419
900	95 424	429	434	439	444	448	453	458	463	468
N	0	1	2	3	4	5	6	7	8	9

Prop. Parts

	6
1	0.6
2	1.2
3	1.8
4	2.4
5	3.0
6	3.6
7	4.2
8	4.8
9	5.4

	5
1	0.5
2	1.0
3	1.5
4	2.0
5	2.5
6	3.0
7	3.5
8	4.0
9	4.5

	4
1	0.4
2	0.8
3	1.2
4	1.6
5	2.0
6	2.4
7	2.8
8	3.2
9	3.6

Prop. Parts

TABLE V—(*Continued*)
FIVE-PLACE LOGARITHMS: 900–950

N	0	1	2	3	4	5	6	7	8	9
900	95 424	429	434	439	444	448	453	458	463	468
01	472	477	482	487	492	497	501	506	511	516
02	521	525	530	535	540	545	550	554	559	564
03	569	574	578	583	588	593	598	602	607	612
04	617	622	626	631	636	641	646	650	655	660
05	665	670	674	679	684	689	694	698	703	708
06	713	718	722	727	732	737	742	746	751	756
07	761	766	770	775	780	785	789	794	799	804
08	809	813	818	823	828	832	837	842	847	852
09	856	861	866	871	875	880	885	890	895	899
910	904	909	914	918	923	928	933	938	942	947
11	952	957	961	966	971	976	980	985	990	995
12	95 999	*004	*009	*014	*019	*023	*028	*033	*038	*042
13	96 047	052	057	061	066	071	076	080	085	090
14	095	099	104	109	114	118	123	128	133	137
15	142	147	152	156	161	166	171	175	180	185
16	190	194	199	204	209	213	218	223	227	232
17	237	242	246	251	256	261	265	270	275	280
18	284	289	294	298	303	308	313	317	322	327
19	332	336	341	346	350	355	360	365	369	374
920	379	384	388	393	398	402	407	412	417	421
21	426	431	435	440	445	450	454	459	464	468
22	473	478	483	487	492	497	501	506	511	515
23	520	525	530	534	539	544	548	553	558	562
24	567	572	577	581	586	591	595	600	605	609
25	614	619	624	628	633	638	642	647	652	656
26	661	666	670	675	680	685	689	694	699	703
27	708	713	717	722	727	731	736	741	745	750
28	755	759	764	769	774	778	783	788	792	797
29	802	806	811	816	820	825	830	834	839	844
930	848	853	858	862	867	872	876	881	886	890
31	895	900	904	909	914	918	923	928	932	937
32	942	946	951	956	960	965	970	974	979	984
33	96 988	993	997	*002	*007	*011	*016	*021	*025	*030
34	97 035	039	044	049	053	058	063	067	072	077
35	081	086	090	095	100	104	109	114	118	123
36	128	132	137	142	146	151	155	160	165	169
37	174	179	183	188	192	197	202	206	211	216
38	220	225	230	234	239	243	248	253	257	262
39	267	271	276	280	285	290	294	299	304	308
940	313	317	322	327	331	336	340	345	350	354
41	359	364	368	373	377	382	387	391	396	400
42	405	410	414	419	424	428	433	437	442	447
43	451	456	460	465	470	474	479	483	488	493
44	497	502	506	511	516	520	525	529	534	539
45	543	548	552	557	562	566	571	575	580	585
46	589	594	598	603	607	612	617	621	626	630
47	635	640	644	649	653	658	663	667	672	676
48	681	685	690	695	699	704	708	713	717	722
49	727	731	736	740	745	749	754	759	763	768
950	97 772	777	782	786	791	795	800	804	809	813
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

Prop. Parts

	5
1	0.5
2	1.0
3	1.5
4	2.0
5	2.5
6	3.0
7	3.5
8	4.0
9	4.5

	4
1	0.4
2	0.8
3	1.2
4	1.6
5	2.0
6	2.4
7	2.8
8	3.2
9	3.6

Prop. Parts

TABLE V—(*Continued*)

FIVE-PLACE LOGARITHMS: 950–1000

Prop. Parts	N	0	1	2	3	4	5	6	7	8	9
	950	97 772	777	782	786	791	795	800	804	809	813
	51	818	823	827	832	836	841	845	850	855	859
	52	864	868	873	877	882	886	891	896	900	905
	53	909	914	918	923	928	932	937	941	946	950
	54	97 955	959	964	968	973	978	982	987	991	996
	55	98 000	005	009	014	019	023	028	032	037	041
	56	046	050	055	059	064	068	073	078	082	087
	57	091	096	100	105	109	114	118	123	127	132
	58	137	141	146	150	155	159	164	168	173	177
	59	182	186	191	195	200	204	209	214	218	223
	960	227	232	236	241	245	250	254	259	263	268
	61	272	277	281	286	290	295	299	304	308	313
5	62	318	322	327	331	336	340	345	349	354	358
1 0.5; **2** 1.0	63	363	367	372	376	381	385	390	394	399	403
3 1.5	64	408	412	417	421	426	430	435	439	444	448
4 2.0	65	453	457	462	466	471	475	480	484	489	493
5 2.5; **6** 3.0	66	498	502	507	511	516	520	525	529	534	538
7 3.5	67	543	547	552	556	561	565	570	574	579	583
8 4.0	68	588	592	597	601	605	610	614	619	623	628
9 4.5	69	632	637	641	646	650	655	659	664	668	673
	970	677	682	686	691	695	700	704	709	713	717
	71	722	726	731	735	740	744	749	753	758	762
	72	767	771	776	780	784	789	793	798	802	807
	73	811	816	820	825	829	834	838	843	847	851
	74	856	860	865	869	874	878	883	887	892	896
	75	900	905	909	914	918	923	927	932	936	941
	76	945	949	954	958	963	967	972	976	981	985
	77	98 989	994	998	*003	*007	*012	*016	*021	*025	*029
	78	99 034	038	043	047	052	056	061	065	069	074
	79	078	083	087	092	096	100	105	109	114	118
	980	123	127	131	136	140	145	149	154	158	162
4	81	167	171	176	180	185	189	193	198	202	207
1 0.4	82	211	216	220	224	229	233	238	242	247	251
2 0.8; **3** 1.2	83	255	260	264	269	273	277	282	286	291	295
4 1.6	84	300	304	308	313	317	322	326	330	335	339
5 2.0	85	344	348	352	357	361	366	370	374	379	383
6 2.4; **7** 2.8	86	388	392	396	401	405	410	414	419	423	427
8 3.2	87	432	436	441	445	449	454	458	463	467	471
9 3.6	88	476	480	484	489	493	498	502	506	511	515
	89	520	524	528	533	537	542	546	550	555	559
	990	564	568	572	577	581	585	590	594	599	603
	91	607	612	616	621	625	629	634	638	642	647
	92	651	656	660	664	669	673	677	682	686	691
	93	695	699	704	708	712	717	721	726	730	734
	94	739	743	747	752	756	760	765	769	774	778
	95	782	787	791	795	800	804	808	813	817	822
	96	826	830	835	839	843	848	852	856	861	865
	97	870	874	878	883	887	891	896	900	904	909
	98	913	917	922	926	930	935	939	944	948	952
	99	99 957	961	965	970	974	978	983	987	991	996
	1000	00 000	004	009	013	017	022	026	030	035	039
Prop. Parts	N	0	1	2	3	4	5	6	7	8	9

TABLE VI*

Natural Logarithms of Numbers

$$\text{Base } e = 2.71828\ldots$$

Note. — $\log_e 10N = \log_e N + \log_e 10$

$$\log_e \frac{N}{10} = \log_e N - \log_e 10$$

$$\log_e 10 = 2.30259$$

Examples. $\log_e 35 = \log_e 3.5 + \log_e 10$
$= 1.25276 + 2.30259 = 3.55535$

$\log_e .35 = \log_e 3.5 - \log_e 10$
$= 1.25276 - 2.30259 = 8.95017 - 10$

N	0	1	2	3	4	5	6	7	8	9
1.0	0.0 0000	0995	1980	2956	3922	4879	5827	6766	7696	8618
1.1	9531	*0436	*1333	*2222	*3103	*3976	*4842	*5700	*6551	*7395
1.2	0.1 8232	9062	9885	*0701	*1511	*2314	*3111	*3902	*4686	*5464
1.3	0.2 6236	7003	7763	8518	9267	*0010	*0748	*1481	*2208	*2930
1.4	0.3 3647	4359	5066	5767	6464	7156	7844	8526	9204	9878
1.5	0.4 0547	1211	1871	2527	3178	3825	4469	5108	5742	6373
1.6	7000	7623	8243	8858	9470	*0078	*0682	*1282	*1879	*2473
1.7	0.5 3063	3649	4232	4812	5389	5962	6531	7098	7661	8222
1.8	8779	9333	9884	*0432	*0977	*1519	*2058	*2594	*3127	*3658
1.9	0.6 4185	4710	5233	5752	6269	6783	7294	7803	8310	8813
2.0	9315	9813	*0310	*0804	*1295	*1784	*2271	*2755	*3237	*3716
2.1	0.7 4194	4669	5142	5612	6081	6547	7011	7473	7932	8390
2.2	8846	9299	9751	*0200	*0648	*1093	*1536	*1978	*2418	*2855
2.3	0.8 3291	3725	4157	4587	5015	5442	5866	6289	6710	7129
2.4	7547	7963	8377	8789	9200	9609	*0016	*0422	*0826	*1228
2.5	0.9 1629	2028	2426	2822	3216	3609	4001	4391	4779	5166
2.6	5551	5935	6317	6698	7078	7456	7833	8208	8582	8954
2.7	9325	9695	*0063	*0430	*0796	*1160	*1523	*1885	*2245	*2604
2.8	1.0 2962	3318	3674	4028	4380	4732	5082	5431	5779	6126
2.9	6471	6815	7158	7500	7841	8181	8519	8856	9192	9527
3.0	9861	*0194	*0526	*0856	*1186	*1514	*1841	*2168	*2493	*2817
3.1	1.1 3140	3462	3783	4103	4422	4740	5057	5373	5688	6002
3.2	6315	6627	6938	7248	7557	7865	8173	8479	8784	9089
3.3	9392	9695	9996	*0297	*0597	*0896	*1194	*1491	*1788	*2083
3.4	1.2 2378	2671	2964	3256	3547	3837	4127	4415	4703	4990
3.5	5276	5562	5846	6130	6413	6695	6976	7257	7536	7815
3.6	8093	8371	8647	8923	9198	9473	9746	*0019	*0291	*0563
3.7	1.3 0833	1103	1372	1641	1909	2176	2442	2708	2972	3237
3.8	3500	3763	4025	4286	4547	4807	5067	5325	5584	5841
3.9	6098	6354	6609	6864	7118	7372	7624	7877	8128	8379
4.0	8629	8879	9128	9377	9624	9872	*0118	*0364	*0610	*0854
4.1	1.4 1099	1342	1585	1828	2070	2311	2552	2792	3031	3270
4.2	3508	3746	3984	4220	4456	4692	4927	5161	5395	5629
4.3	5862	6094	6326	6557	6787	7018	7247	7476	7705	7933
4.4	8160	8387	8614	8840	9065	9290	9515	9739	9962	*0185
4.5	1.5 0408	0630	0851	1072	1293	1513	1732	1951	2170	2388
4.6	2606	2823	3039	3256	3471	3687	3902	4116	4330	4543
4.7	4756	4969	5181	5393	5604	5814	6025	6235	6444	6653
4.8	6862	7070	7277	7485	7691	7898	8104	8309	8515	8719
4.9	8924	9127	9331	9534	9737	9939	*0141	*0342	*0543	*0744
5.0	1.6 0944	1144	1343	1542	1741	1939	2137	2334	2531	2728
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

* This table was taken from "Plane Trigonometry with Five-place Tables" by H. A. Simmons and G. D. Gore by permission of the authors and the publisher, John Wiley & Sons, Inc.

TABLE VI—(*Continued*)

FIVE-PLACE NATURAL LOGARITHMS

N	0	1	2	3	4	5	6	7	8	9
5.0	1.6 0944	1144	1343	1542	1741	1939	2137	2334	2531	2728
5.1	2924	3120	3315	3511	3705	3900	4094	4287	4481	4673
5.2	4866	5058	5250	5441	5632	5823	6013	6203	6393	6582
5.3	6771	6959	7147	7335	7523	7710	7896	8083	8269	8455
5.4	8640	8825	9010	9194	9378	9562	9745	9928	*0111	*0293
5.5	1.7 0475	0656	0838	1019	1199	1380	1560	1740	1919	2098
5.6	2277	2455	2633	2811	2988	3166	3342	3519	3695	3871
5.7	4047	4222	4397	4572	4746	4920	5094	5267	5440	5613
5.8	5786	5958	6130	6302	6473	6644	6815	6985	7156	7326
5.9	7495	7665	7834	8002	8171	8339	8507	8675	8842	9009
6.0	9176	9342	9509	9675	9840	*0006	*0171	*0336	*0500	*0665
6.1	1.8 0829	0993	1156	1319	1482	1645	1808	1970	2132	2294
6.2	2455	2616	2777	2938	3098	3258	3418	3578	3737	3896
6.3	4055	4214	4372	4530	4688	4845	5003	5160	5317	5473
6.4	5630	5786	5942	6097	6253	6408	6563	6718	6872	7026
6.5	7180	7334	7487	7641	7794	7947	8099	8251	8403	8555
6.6	8707	8858	9010	9160	9311	9462	9612	9762	9912	*0061
6.7	1.9 0211	0360	0509	0658	0806	0954	1102	1250	1398	1545
6.8	1692	1839	1986	2132	2279	2425	2571	2716	2862	3007
6.9	3152	3297	3442	3586	3730	3874	4018	4162	4305	4448
7.0	4591	4734	4876	5019	5161	5303	5445	5586	5727	5869
7.1	6009	6150	6291	6431	6571	6711	6851	6991	7130	7269
7.2	7408	7547	7685	7824	7962	8100	8238	8376	8513	8650
7.3	8787	8924	9061	9198	9334	9470	9606	9742	9877	*0013
7.4	2.0 0148	0283	0418	0553	0687	0821	0956	1089	1223	1357
7.5	1490	1624	1757	1890	2022	2155	2287	2419	2551	2683
7.6	2815	2946	3078	3209	3340	3471	3601	3732	3862	3992
7.7	4122	4252	4381	4511	4640	4769	4898	5027	5156	5284
7.8	5412	5540	5668	5796	5924	6051	6179	6306	6433	6560
7.9	6686	6813	6939	7065	7191	7317	7443	7568	7694	7819
8.0	7944	8069	8194	8318	8443	8567	8691	8815	8939	9063
8.1	9186	9310	9433	9556	9679	9802	9924	*0047	*0169	*0291
8.2	2.1 0413	0535	0657	0779	0900	1021	1142	1263	1384	1505
8.3	1626	1746	1866	1986	2106	2226	2346	2465	2585	2704
8.4	2823	2942	3061	3180	3298	3417	3535	3653	3771	3889
8.5	4007	4124	4242	4359	4476	4593	4710	4827	4943	5060
8.6	5176	5292	5409	5524	5640	5756	5871	5987	6102	6217
8.7	6332	6447	6562	6677	6791	6905	7020	7134	7248	7361
8.8	7475	7589	7702	7816	7929	8042	8155	8267	8380	8493
8.9	8605	8717	8830	8942	9054	9165	9277	9389	9500	9611
9.0	9722	9834	9944	*0055	*0166	*0276	*0387	*0497	*0607	*0717
9.1	2.2 0827	0937	1047	1157	1266	1375	1485	1594	1703	1812
9.2	1920	2029	2138	2246	2354	2462	2570	2678	2786	2894
9.3	3001	3109	3216	3324	3431	3538	3645	3751	3858	3965
9.4	4071	4177	4284	4390	4496	4601	4707	4813	4918	5024
9.5	5129	5234	5339	5444	5549	5654	5759	5863	5968	6072
9.6	6176	6280	6384	6488	6592	6696	6799	6903	7006	7109
9.7	7213	7316	7419	7521	7624	7727	7829	7932	8034	8136
9.8	8238	8340	8442	8544	8646	8747	8849	8950	9051	9152
9.9	9253	9354	9455	9556	9657	9757	9858	9958	*0058	*0158
10.0	2.3 0259	0358	0458	0558	0658	0757	0857	0956	1055	1154
N	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**

ANSWERS

PAGES 4 AND 5

1. 784. **2.** 322. **4.** 9.87269. **5.** $\frac{511}{256}$. **6.** -65. **10.** 120.
13. $n(n+1)(4n-1)/6$. **14.** $2n^2(n+1)$. **15.** $n(n+1)(n+2)(n+3)/4$.
16. $2n^2(n^2-1)$. **17.** 2,485. **18.** $\sum_{x=7}^{x=15} x(x+3)$. **19.** $\sum_{15}^{42} vf(v)$.
20. $\sum_{4}^{12} v^3(v+1)$. **21.** $\sum_{i=10}^{18} \frac{1}{V_i}$. **22.** $\sum_{i=6}^{30} v_i f(v_i)$.
23. $\sum_{7}^{11} x(3x^2+5x-4x)$.

PAGES 22 AND 23

1. 11 seeds. **2.** (*a*) 2.16 in. (*b*) 2.17 in. **3.** 461.2 gal.
4. (*a*) 204.35 eggs. (*b*) 5.11 eggs.

PAGES 24 AND 25

1. 17.32 ligulate flowers. **2.** 200.27 loaves.
3. (*a*) 73.62. (*b*) Per cent A's = 0.93, per cent of C's = 47.56.

PAGE 27

1. 146 students.

PAGES 28 AND 29

1. 9 312 petals. **2.** 88.45 strokes. **3.** 24,755.7 lb. **4.** 25 pupils.
5. 40.075 qt. **6.** 12.76 rays.

PAGES 29 AND 30

1. 75.7. **2.** 139.51 lb. **3.** (*a*) 4,804 gal. per day. (*b*) 148,924. **6.** \$199.70.

PAGES 35 AND 36

1. 79.33. **2.** 17.088 rays. **3.** 171.82 cu. in. **4.** 120.02 lb.
5. (*a*) 53.51. (*b*) 53.26. **6.** \$5.75.

PAGE 38

4. 0.119.

PAGE 40

4. (*a*) 53. (*b*) 20. **5.** (*a*) 215.

PAGE 41

1. 79 grade pt. **2.** 22.5 years.

Pages 44 and 45

1. 74.08. Area to right of median is 497 **2.** 25 96 years of age. **3.** $4.80

Page 46

1. 11.06. **2.** 17.14 ft. **3.** 2 92 mice

Pages 48 and 49

1. 0 0047. **2.** (*a*) rate = 0 01425 (*b*) r = 0 00509
3. (*a*) 0.1767. (*b*) 0.1042 **4.** 0 1946. **5.** (*a*) 0 0291 (*b*) 0.2419. **6.** 1.710.
7. 10.

Pages 51 and 52

1. 121.78 mi. per hr. **2.** (*a*) 12 cents per lb (*b*) $8\frac{1}{3}$.
3. (*a*) 9.92 (*b*) 49.68. **4.** $1.75 **5.** (*a*) 0 65. (*b*) about 2 men
7. 4. **8.** 0.04.

Pages 58 and 59

1. (*a*) 20.6 tires. (*b*) 8 including 18. **2.** 1.96 ker. **3.** (*a*) 1 77 Al. Part (*b*) 47 per cent. (*c*) 88 per cent.

Pages 65 and 66

1. (*a*) M = 10.108 sheets, (*b*) S. D. = 5.804 sheets. **2.** 69, 98, 100.
3. 573 or 69.4 per cent, 947 or 98 24 per cent, 964 or 100 per cent. **4.** 101 26.
5. 4.94. **7.** 635,104. **8.** 5 07. **9.** n = 500.

Pages 72 and 73

1. M = 17.52, S. D. = 2.65, 63 per cent, 98.7 per cent, 99.6 per cent.
2. M = 74.2, S. D = 11 067. **3.** M = 74 05, S. D. = 11.074.
4. M = 170.121, S D = 29.74, 67.2 per cent., 94.7 per cent, 99 8 per cent.
(*b*) Prob. = 0.116.

Pages 76 and 77

1. M = 66.53, S. D. = 2.96. **2.** M = 66.67, S. D. = 3.02.
4. M = 72.1, S. D. = 4.6. **5.** M = 138.92, S. D. = 16 8.
6. M = 17.24, S. D. = 1.17.

Pages 80 and 81

1. 178. **2.** S. D. = 12. **3.** M = 132.86, S. D. = 14.17 lb., V = 10.7
(*b*) M = 228.66 lb , S. D. = 16.88 lb., V = 7.4. **4.** Q = 8.05; 395.
5. V = 15.00, V = 14 97.

7. $P_{83} = C_{83} + \dfrac{(\frac{83}{100}\, n - n_{83})}{f_{83}} W_{83}$. **9.** $D_7 = C_7 + \dfrac{(\frac{7}{10}\, n - n_7)}{f_7} W_7$.

Pages 89 and 90

1. Exp. freq. 3, 8, 16, 29, 45, 59, 65, 61, 48, 33, 19, 9, 4, 1; sum = 400.
2. A's = 11, B's = D's = 71, C's = 136, E's = 11; sum = 300.
3. A's = E's = 1.77, B's = D's = 22.41, C's = 51.60; sum = 99.96.
4. $A_+ = E_-$ = 0.33, A = E = 0.92, $A_- = E_+$ = 2.20, $B_+ = D_-$ = 4.48, B = D = 7.79. $B_- = D_+$ = 11.56, $C_+ = C_-$ = 14.65, C = 15.85, sum = 99.71. **5.** Exp. freq. 4, 11, 27, 52, 84, 109, 123, 112, 85, 53, 27, 12, 4; sum = 703; 2 without this range. **7.** V = 98.28. **9.** 216.56 gal. = M.
10. 4.44 cc.

Page 95

1. Exp. freq. 1, 3, 7, 13, 18, 20, 18, 12, 6, 2, 1; sum = 101. **2.** $t_1 = 0\ 80, t_2 = 1.32$; area between = 0 118437, N = 2,930 **3.** N = 5,561. **4.** N = 3,897. **5.** h = 101. **6.** 7's = 11's = 48, $7\frac{1}{2} = 10\frac{1}{2}$ = 404, 8's = 10's = 1,846, $8\frac{1}{2} = 9\frac{1}{2}$ = 4,590, 9's = 6,217.

Pages 102 and 103

1. 1,010.6 = A, B = 2,642.6, C = 2,886.60, D = 719.4, E = 225 0. **2.** S D = 2 cm. **3.** 1,721.6. **4.** N = 500 **5.** 53 67 rays. S. D. = 2.131 rays. Skew. = −.079. **6.** M = 24.87 yr , S. D = 7.8 yr. Skew. = 1 8.

$$7.\ \alpha_{3:x+y} = \frac{1}{\sigma^3_{x+y}}\left\{\frac{n_1[\sigma_x{}^3\alpha_{3\cdot x}+3M_x\ \sigma_x{}^2+M_x{}^2]+n_2[\sigma_y{}^3\ \alpha_{3\ y}+3M_y\ \sigma_y{}^2+M_y{}^3]}{n_1+n_2} - 3M_{x+y}\sigma_{x+y}{}^2 - M_{x+y}{}^3\right\}.$$

8. M_{x+y} = 139.03 lb , S. D$_{x+y}$ = 17.89 lb., Skew.$_{x+y}$ = 0.92.

Page 107

1. Skew. = −0.077, K = 1.10 **2.** M = 12.804 stamen, S D. = 1.4407 stamen, Skew. = .644, K = 0.464. **3.** α_4 = 0.88.

Pages 108 and 109

1. 15. **2.** 216. **3.** 24. **4.** 25. **5.** (*a*) 9, (*b*) 9. **6.** 49. **7.** 60. **8.** 90. **9.** 5^5.

Page 110

1. 504. **2.** 648. **3.** 60,480. **4.** 10! **5.** 120. **6.** 3,628,800. **7.** 4,320 **9.** 4.

Page 113

1. 55. **2.** 10. **3.** 700. **4.** 360. **5.** (*a*) 210, (*b*) 214. **6.** (*a*) 60, (*b*) 21, (*c*) 480, (*d*) 1, (*e*) 630. **7.** 3,150. **8.** 210. **9.** ${}_{52}C_{13}$ **10.** 132. **11.** 63. **12.** 52.

Pages 114 and 115

1. (*a*) $\frac{4}{36}$, (*b*) $\frac{2}{36}$, (*c*) $\frac{3}{36}$, (*d*) $\frac{6}{36}$.

2. (*a*) $\frac{10}{84}$, (*b*) $\frac{80}{84}$, (*c*) 0, (*d*) $\frac{34}{84}$, (*e*) $\frac{30}{84}$.

3. $\frac{1}{24}$, $\frac{4}{24}$, $\frac{4}{24}$, $\frac{1}{24}$ **4.** (*a*) 0 0032, (*b*) 0 0368, (*c*) 0.8832, (*d*) $1 - (\frac{11}{12})^5$, (*e*) $(\frac{1}{12})^5$. **5.** $1 - (\frac{23}{24})^5$, (*b*) $(\frac{1}{24})^5$, (*c*) $(\frac{11}{12})^5$, (*d*) $1 - (\frac{11}{12})^5$. **6.** (*a*) 0 374, (*b*) 0.047, (*c*) 0.2299, (*d*) 0.08199 **7.** 0 001.

Pages 118 and 119

1. $\frac{1}{4096}$, (*b*) $\frac{694}{4096}$, (*c*) $\frac{1}{4096}$, (*d*) $\frac{3402}{4096}$, (*e*) $\frac{693}{4096}$
2. (*a*) $\frac{560}{2187}$, (*b*) $\frac{1120}{2187}$, (*c*) $\frac{99}{2187}$, (*d*) $\frac{84}{2187}$, (*e*) $\frac{1}{2187}$, (*f*) $\frac{939}{2187}$
3. (*a*) $\frac{45}{1024}$, (*b*) $\frac{210}{1024}$, (*c*) $\frac{56}{1024}$, (*d*) $\frac{10}{1024}$
4. (*a*) $6(0.98)^5(0.02)$, (*b*) $(0.98)^6+6(0.98)^5(0.02)+15(0.98)^4(0.02)^2$, (*c*) $(0\ 98)^6$. **5.** (*a*) $(0.92)^3$, (*b*) 0.203136, (*c*) 0.067712, (*d*) 0.000512. **6.** (*a*) $\frac{216}{625}$, (*b*) $\frac{96}{625}$, (*c*) $\frac{16}{625}$, (*d*) $\frac{513}{625}$. **7.** (*a*) 0.0446, (*b*) 0.3050. (*c*) 0.78642, (*d*) 0.0307, (*e*) 0.02951, (*f*) 0.1095^-. **8.** w = 20.23. **9.** (*a*) 0.6828, (*b*) 0.0214. **10.** (*a*) 0.50, (*b*) 0.25.

Pages 128 and 129

1. $p = 0.02076$. **2.** (a) $\frac{252}{210}$ (b) $\frac{120}{210}$ **3.** (a) 0 051551, (b) 0 09680, (c) 0.851649, (d) 0.026594, (e) ${}_{1200}C_{300} \cdot (\frac{1}{4})^{300}(\frac{3}{4})^{900}$. **4.** (a) 0 351973, (b) 0.344578, (c) 0 03128 **5.** $M_{\text{new}} = H\,M_{\text{old}}$, $\text{S D}._{\text{new}} = H(\text{S D}_{\text{old}})$, $\text{Skew}._{\text{new}} = \text{Skew}_{\text{old}}$. **6.** $K = (1 - 6pq)/npq$ (a) Zero.

Pages 142 and 143

5. Ideal index for 1934 = 0.734.

Page 144

1. (b)

1926,	1927,	1928,	1929,	1930.
51,	52,	53,	63,	100.

2. (a)

1930,	1931,	1932,	1933,	1934,	1935,	1936,	1937.
112,	101,	85,	90,	100,	107,	110,	117.

Page 147

1. (b) min. $= -11\frac{1}{8}$. **2.** $x = 4.2$, $y = 5\,2$, $\Sigma e^2 = 0.05$.

Pages 149 and 150

1. $x = 3\,71$, $y = 7.72$, $\Sigma e^2 = 0\,0014$. **2.** $x = 4\,8$, $y = 2.2$, $z = 7.3$, $e_1 = 0.1$, $e_3 = -0.2$. **3.** $x = 1$, $y = 2$. **4.** 10. **6.** $AB = 7.3$ mi., $BC = 11.3$ mi.

Pages 153 and 154

1. $y = -1.695^{+} + 1.646\,x$, $\sigma_e = 7.30$ in. **2.** Between 11.5 and 15.7, may be between 10 and 16. **3.** Zero.

Page 155

1. $y = 11.122 + 0.732\,x$, $\sigma_e = 0.287$. **2.** $n = 153$.

Page 159

1. $y = -137,\,017.82 + 72.408\,x$, $\sigma_e = 115.9$ pupils.

Pages 162 and 163

1. $y = 3.089 + 0\,976\,x$, $\sigma_e = 0.456$ cm. **2.** $y = 38.6 + 1.57\,x$, $\sigma_e = 3.52$, (b) averages.

Page 165

2. $y = -272.7 + 2.633\,x + 6.653\,z$, $\sigma_e = 8.452$ lb. **3.** $y = -106.654 + 3.109\,x + 7.144\,z$, $\sigma_e = 5.63$ lb.

Page 167

1. $y = -197.21 + 1.388\,x + 3.247\,z + 6.119\,w$, $\sigma_e = 4.56$ lb.

8.
$$\begin{aligned} na + \Sigma xb + \Sigma x^2 z &= \Sigma y, \\ \Sigma xa + \Sigma x^2 b + \Sigma x^3 c &= \Sigma xy, \\ \Sigma x^2 a + \Sigma x^3 b + \Sigma x^4 c &= \Sigma x^2 y. \end{aligned}$$

9. $\Sigma(a/x) = \Sigma y$.

Pages 171 and 172

1. $r = 0.568$. **2.** $r = 0.807$. **4.** $r = 0.707$. **5.** When the predicting line is parallel to horizontal axis and all points fall on the line.

PAGES 175 AND 176

1. $R = .681$. **2.** $R = 0.751$.

PAGE 177

1. $x = 56.014 + 0\,0814\ y$, $\theta = 29°\ 35'$, $B = 30°\ 50'$.

PAGES 182 AND 183

1. $r = 0.622$. **2.** $r = 0.788$.

PAGES 185 AND 186

1. $y = 66.88 + 0.0121\ x$. **2.** $\bar{y} = 15.59$. **3.** $y = -217.7 + 11.99\ x$, $y = 152.8$ when $x = 30.9$. **4.** $\sigma_y = 25$ lb.

PAGES 188 AND 189

1. $r = \pm\sqrt{0.99}$. **2.** (*a*) $R_{y\ xz} = 0$. **4.** $R_{y.xz} = 0.885^-$. **5.** $R_{y.xz} = 0.849$. **8.** $R = \sqrt{0.986}$. **9.** $\bar{y} = \frac{\sigma_y}{\sigma_x}\left[\frac{r_{xy} - r_{xz}r_{yz}}{1 - r_{xz}^2}\right]\bar{x} + \frac{\sigma_y}{\sigma_z}\left[\frac{r_{zy} - r_{xz}r_{xy}}{1 - r_{xz}^2}\right]\bar{z}$.
10. $R_{y.xz} = 0.9992$.

PAGES 192 AND 193

1. $D_0 = 0.42$, $R_{y.xzw} = 0.97$. **4.** $D_0 = 0.3039$, $D_1 = 0.1001$, $D_3 = 0.2043$, $D_2 = 0.0326$. $\bar{y} = 2.382\ \bar{x}_1 + 1.024\ \bar{x}_2 + 3.627\ \bar{x}_3$, $R_{y.x_1\,x_2\,x_3} = 0.977$.

PAGES 196 AND 197

1. $r_{me\ a} = -0.0247$. **2.** $r_{wh} = 0\,566$, $r_{wt} = 0.919$, $r_{ht} = 0.448$, $r_{wt.h} = 0.437$. **3.** $r_{me} = 0.53$, $r_{m\ 4\ \text{yr}} = 0.74$, $r_{e\ 4\ \text{yr.}} = 0.73$, $r_{me\cdot\ 4\ \text{yr.}} = -0.022$. **14.** Zero.

PAGES 199 AND 200

1. $r = 0.23$. **2.** $r = -.493$. **3.** $r = -.869$.

PAGE 205

2. (*a*) 210, (*b*) 210, (*c*) 252.

PAGE 208

2. (*a*) = 0.159, (*b*) 0.08, (*c*) 0.04.

PAGES 210 AND 211

1. $M_{400} = 36.8$ in., S.D. = 0 046, Skew. = 0 008. **2.** S.D. = 0.054, Skew. = 0.013. **3.** (*a*) 0.000032, (*b*) 0.022216, (*c*) = 1 −. **4.** $t = 5$, not due to random sampling. **6.** $s > 4{,}375$. **10.** ${}_{100}C_5$, ${}_{100}C_{25}$. **11.** (*a*) $r > 14$, (*b*) $r > 54$, (*c*) $r > 214$. **12.** (*a*) $r = 4$, (*b*) $r = 16$, (*c*) $r = 64$, (*d*) $r = 256$, (*e*) $r = 25600$. **13.** $r = 121$.

PAGE 216

2. $n = 100$. **3.** 0.56. **4.** (*a*) 2, (*b*) $2/\sqrt{37}$, (*c*) $\sqrt{144/37}$.

PAGES 218 AND 219

1. (*a*) $AF = 37.8 \pm 0.44$, (*b*) S.D. = 0.12. **2.** 2.53. **3.** $AB = 1.44 \pm 04$.

Pages 221 and 222

1. $t > 6$, sign. **2.** $t > 2\ 6$, sign. **3.** $n > 78$ trees. **4.** No danger. **5.** $t = 1.718/0\ 218 > 2.6$; sign **6.** $t > 2.6$; sign.

Pages 225 and 226

3. $y = -22.595 + 10.2944\,x + 0.1342\,x^2$. $\sigma_e = 4.8$. $r = 0.989$.

Pages 231 and 232

1. $r_{yx} = 0.707$, $\eta_{yx} = 0.792$. $\eta_{xy} = 0\ 784$ **2.** 0.857.

Page 236

1. $y = 398.976 - .201\,x$.

Pages 261 and 263

1. Analysis of Variances

Source of Variation	Degrees of Freedom	Sum of Squares	Variances	Exp. Error
Total	24	1,977 84		
Col.	4	61 04	15.26	
Rows	4	76 24	19 06	
Treatments	4	1,737 84	434 46	
Error	12	102.72	9 56	3 09

Treatments M and N significantly better than the others.

2. Analyses of Variances

Source of Variation	Degrees of Freedom	Sum of Squares	Variances	Exp. Error
Total	59	16 27		
Between means	2	2.21	1 11	
Within states	57	14 06	0.25	0 5

State C better than State B.

3. Variety B is better than the others.

Page 243

July 1, 1930, 2.30	July 1, 1931, 1.77	July 1, 1932, 1.32	July 1, 1933, 1.30	July 1, 1934, 1.53

PAGES 269, 270 AND 271

1. Area $= 1{,}211.74 \pm 5.8$ sq. ft. **2.** Cir. $= 185.354 \pm .628$. **3.** Vol. $= 8{,}158\ 74 \pm 33.96$. **4.** 29 7 sq. ft. **5.** 15.4. **6.** $137{,}632 \pm 1{,}481.2$ cu. ft **7.** $t = 0.002/0.019$; not sign **9.** $V = 14\ 97$, $V = 14.95$, not sign. **10.** $t = .21/.18$; not sign **11.** $\sigma_r = 0\ 0198$ **13.** $t = 0\ 08/0\ 03$; sign. **16.** $b = 12.79 \pm .051$ **18.** $\sigma_{v_1} = 0.00028$, $\sigma_{v_2} = 0\ 00033$; $t = 0\ 0051/0.0004$; sign. **19.** Yes.

PAGES 278 AND 279

1. $t = 1\ 83$; not sign. **2.** April 22 better than other dates. **3.** Diet B is the better.

PAGE 280

1. Not significantly different from zero.

INDEX

(Numbers refer to pages)

w

2023